University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Theses, Dissertations, and Student Research in Agronomy and Horticulture

Agronomy and Horticulture Department

7-2017

Dissecting the Genetic Basis of Salt Tolerance in Rice (Oryza sativa)

Malachy Thomas Campbell
University of Nebraska-Lincoln

Follow this and additional works at: http://digitalcommons.unl.edu/agronhortdiss

 Part of the Plant Biology Commons, and the Plant Breeding and Genetics Commons

Campbell, Malachy Thomas, "Dissecting the Genetic Basis of Salt Tolerance in Rice (Oryza sativa)" (2017). *Theses, Dissertations, and Student Research in Agronomy and Horticulture*. 127.
http://digitalcommons.unl.edu/agronhortdiss/127

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Theses, Dissertations, and Student Research in Agronomy and Horticulture by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

DISSECTING THE GENETIC BASIS OF SALT TOLERANCE IN RICE (*Oryza sativa*)

by

Malachy Thomas Campbell

A DISSERTATION

Presented to the faculty of

The Graduate College of

The University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Agronomy and Horticulture

(Plant Breeding and Genetics)

Under the Supervision of Professor Harkamal Walia

Lincoln, Nebraska

July, 2017

DISSECTING THE GENETIC BASIS OF SALT TOLERANCE IN RICE (*Oryza sativa*)

Malachy Thomas Campbell, Ph.D.

University of Nebraska, 2017

Advisor: Harkamal Walia

Excessive salt is a major factor limiting crop productivity. Rice is an important food crop worldwide, and provides a major source of calories for billions of people in tropical and subtropical developing countries. In many regions throughout Southeast Asia, rice is grown in low-lying coastal regions, which are prone to ingress by seawater. Rice is highly sensitive to salt stress, with mild salinity resulting in large reductions in yield. Improving salt tolerance in rice is essential for ensuring food security for many individuals throughput the developing world.

In this work, the overall goals were to characterize a diverse collection of cultivated rice for salt tolerance, and elucidate the genetic basis of salt tolerance-related traits. To this end, a multidisciplinary approach that combined aspects of classical plant physiology, molecular biology, computational biology, and quantitative genetics was employed. Considerable natural variation was observed for salt-tolerance related traits. Moreover, the five subpopulations of rice (*indica*, *temperate japonica*, *tropical japonica*, *aromatic*, and *aus*) exhibited contrasting phenotypes, suggesting that there may be inherent differences between the subpopulations of cultivated rice for salt tolerance.

Phenotypes derived from classical end-point phenotyping, and temporal non-destructive imaging, were used for genome-wide association mapping. For the majority of the traits analyzed, numerous loci were identified that accounted for only a small portion of the phenotypic variation. These results indicate that many genes collectively contribute to salt-tolerance in rice. However, a single gene, *HKT1;1*, was identified and was shown to regulate root

sodium content. Moreover, variants of HKT1;1 were shown to underlie the divergence of *Indica* and *Japonica* subspecies of rice for root Na^+ content. Transcriptome data from 91 diverse accessions of the rice diversity panel were leveraged to elucidate the genetic mechanisms underlying the complex salt tolerance-related traits. Together, this work provides a comprehensive phenotypic and genetic analysis of the responses to moderate salinity.

DEDICATION

I would like to dedicate this work to my family (Maureen, Dominic, Conor, Aedan, Rory, Lisa, Shane, and Allanah) for supporting my many endeavors, my grandmothers Alice Campbell and Eleanor Quinlan, and to Drs. Gojko Jelenkovic and Richard Merritt for cultivating my curiosity in plant science and research.

Acknowledgments

I would like to thank my advisor Harkamal Walia, for giving me the opportunity to develop and test my own hypotheses, and most importantly his patience. I would also like to express my gratitude to the members of my committee: Drs. Paul Staswick, Gautam Sarath, and Chi Zhang; my lab members: Dr. Jing Jin Folsom, Dr. Kevin Begcy Padilla, Dr. Dante Placido, Larissa Irvin, Dr. Chen Chen, Martha Rowe, Dr. Sandeep Sharma, Dr. Alex Grondin, Harel Bacher, Jaspinder Singh, Dr. Puneet Paul, Jaspreet Sandhu, Manny Saluja, Dr. Sajag Adhikari, Balpreet Dhatt, and most importantly Dr. Aaron Schmitz; collaborators: Dr. Aaron Lorenz, Nonoy Bandillo, Dr. Anne-Aliénor Véry, Dr. Bettina Berger, Qian Du, Dr. Kan Lui, and Dr. Argelia Lorence; and my loving girlfriend Christina Hein.

Table of Contents

CHAPTER 1: INTRODUCTION **12**

1.1 Physiological responses to salinity **12**

1.1.1 Osmotic effects of salinity 13

1.1.2 Ionic effects of salinity 14

1.2 Na^+ transport processes in roots **15**

1.3 Na^+ translocation within the plant **18**

1.4 Current challenges and future directions for improving salt tolerance **21**

1.4.1 Assessing the temporal nature of salt response 21

1.4.2 Bridging the genotype-to-phenotype gap 22

1.5 Aims of the current study **25**

References **26**

1.6 Figures **31**

CHAPTER 2: ALLELIC VARIANTS OF *OSHKT1;1* UNDERLIE THE DIVERGENCE BETWEEN *INDICA* AND *JAPONICA* SUBSPECIES OF RICE (*ORYZA SATIVA*) FOR ROOT SODIUM CONTENT **32**

2.1 Introduction **33**

2.2 Results **36**

2.2.1 Rice subpopulations exhibit inherent differences in ion homeostasis 36

2.2.2 Genome-wide association mapping identifies a major QTL on chromosome 4 for root Na^+ and Na^+:K^+ 37

2.2.3 HKT1;1 Regulates Na^+ homeostasis during the early tillering stage 39

2.2.4 Allelic variants of HKT1;1 alter Na^+ transport 41

2.2.5 Origins of HKT1;1 variants in cultivated rice 44

2.3 Discussion 45

2.3.1 HKT1;1 regulates root Na^+ content 47

2.3.2 Variants in HKT1;1 underlie natural variation for root Na^+ content 48

2.3.3 HKT1;1 isoforms are derived from independent populations during domestication 48

2.3.4 The role of HKT1;1 in Na^+ exclusion from shoot tissue 50

2.4 Materials and Methods 53

2.4.1 Plant materials and genotyping 53

2.4.2 Growth conditions and salt treatment 53

2.4.3 Ion content measurement 54

2.4.4 Statistical analysis of phenotypic data 54

2.4.5 Mixed linear model for genome-wide association analysis 55

2.4.6 Estimation of genetic correlation 57

2.4.7 Haplotype analysis 57

2.4.8 Haplotype analysis of HKT1;1 in cultivated rice and Oryza rufipogon 58

2.4.9 Growth conditions and sample collection for RNA sequencing 58

2.4.10 RNA-seq mapping and quantification 59

2.4.11 Variant identification 59

2.4.12 First strand cDNA synthesis and real-time quantitative PCR 60

2.4.13 Transgene construction and Agrobacterium-mediated transformation 60

2.4.14 Evaluation of transgenic plants 61

2.4.15 Electrophysiology 62

2.5 References 64

2.6 Figures 71

2.7 Tables 81

2.8 Supplemental Materials 83

CHAPTER 3: IMAGE HARVEST: AN OPEN SOURCE PLATFORM FOR HIGH-THROUGHPUT PLANT IMAGE PROCESSING AND ANALYSIS 106

3.1 Introduction 107

3.2 Results 110

3.2.1 Software Overview 110

3.2.2 Image Capture 111

3.2.3 Developing a processing workflow on a local computer 112

3.2.4 Executing processing workflows on a computing cluster 114

3.2.5 Assessing plant growth with IH 118

3.2.6 From pixel-based digital traits to genes: A case study in rice 119

3.3 Discussion 120

3.4 Materials and Methods 125

3.4.1 Plant materials 125

3.4.2 Plant Growth Conditions 125

3.4.3 Phenotypic data analysis 126

3.4.4 Genome-wide association mapping 127

3.5 Availability and requirements 127

3.6 References 129

3.7 Figures 135

3.8 Tables 142

3.9 Supplemental Materials 143

CHAPTER 4: A COMPREHENSIVE IMAGE-BASED PHENOMICS ANALYSIS REVEALS THE COMPLEX GENETIC ARCHITECTURE OF SHOOT GROWTH DYNAMICS IN RICE (*ORYZA SATIVA*) 155

4.1 Introduction 156

4.2 Results 158

4.2.1 Considerable natural variation for shoot growth dynamics is present in Rice Diversity Panel 1 158

4.2.2 Genetic basis of shoot biomass 160

4.2.3 Functional GWAS analysis 161

4.2.4 Shoot growth dynamics during early tillering may be regulated by gibberellic acid 163

4.2.5 Prediction of shoot growth dynamics and PSA 164

4.3 Discussion 165

4.4 Materials and Methods 168

4.4.1 Plant Materials and genotyping 168

4.4.2 Greenhouse growth conditions and phenotyping 169

4.4.3 Image Processing 169

4.4.4 Statistical Analysis of projected shoot area (PSA) 170

4.4.5 Functional modeling of temporal trends in PSA 171

4.4.6 Genome-wide association analysis of PSA 171

4.4.7 Genomic Selection 173

4.4.8 Hormone Quantification 174

4.4.9 Growth conditions for transcriptome experiment 175

4.4.10 RNA-seq mapping and analysis 175

4.5 References 177

4.6 Figures 182

4.7 Tables **190**

4.8 Supplemental Materials **192**

CHAPTER 5: INTEGRATING IMAGE-BASED PHENOMICS AND ASSOCIATION ANALYSIS TO DISSECT THE GENETIC ARCHITECTURE OF TEMPORAL SALINITY RESPONSES IN RICE **215**

5.1 Introduction **216**

5.2 Results **219**

5.2.1 Capturing osmotic and ionic components of salinity stress using high-throughput imaging . 219

5.2.2 Assessing salinity-induced growth responses in the rice diversity panel 220

5.2.3 Assessing salinity-induced chlorophyll responses 222

5.2.4 Development of model for association analysis of longitudinal salinity-induced growth responses 223

5.2.5 Genetic architecture of salinity-induced temporal growth response 224

5.2.6 Genetic architecture of salinity-induced chlorophyll responses 226

5.3 Discussion **227**

5.4 Materials and Methods **231**

5.4.1 Greenhouse Conditions and salt treatment 231

5.4.2 VIS/RGB image acquisition and processing 232

5.4.3 Fluorescence image acquisition and processing 233

5.4.4 Determination of sodium and potassium content 235

5.4.5 Hierarchical clustering analysis of fluorescence color classes 235

5.4.6 Canonical correlation analysis 236

5.4.7 Comparison of salinity induced growth response models between subpopulations 236

5.4.8 Longitudinal salinity-induced growth responses association analysis 237

5.4.9 Conventional mixed model genome-wide association analysis 238

5.5 References **239**

5.6 Figures **246**

5.7 Tables **254**

5.8.1 Supplemental Methods: Extended mixed model for association mapping with growth response curve 255

5.8.2 Statistical analysis of phenotypic data 257

CHAPTER 6: TRANSCRIPTIONAL AND REGULATORY VARIATION IN SHOOT TISSUE UNDER SHORT-TERM SALINITY STRESS **302**

6.1 Introduction **302**

6.2 Results and discussion **305**

6.2.1 Differential expression analysis 306

6.2.2 eQTL 308

6.2.3 A transcriptome-wide association study identifies genes potentially underlying complex salt-tolerance related traits 308

6.3 Materials and Methods **310**

6.3.1 Growth conditions for GWAS studies 310

6.3.2 Growth conditions for RNA sequencing study 311

6.3.3 RNA-seq mapping and analysis 312

6.3.4 Microarray processing and analysis 312

6.3.5 Cis-eQTL discovery 313

6.3.6 Transcriptome-wide association study (TWAS) 314

6.3.7 Gene set enrichment analysis (GSEA) 314

6.5 Figures **318**

6.6 Tables **320**

CHAPTER 7: CONCLUDING REMARKS 323

7.1 References..... 325

CHAPTER 1

INTRODUCTION

Soils that contain excessive amounts of sodium are wide spread globally, and significantly reduce the productivity of many cereals. Rice (*Oryza sativa* L.) is arguably the most important crop species and is a staple food for billions of individuals in developing nations. Salinity remains a major constraint on rice production in temperate and tropical environments (Flowers and Yeo, 1995; Chinnusamy et al., 2005).

Among cereals, rice is the most salt-sensitive species (Munns and Tester, 2008). Rice yields may be reduced by 12 percent for every unit (dS m^{-1}) increase in salinity (Redfern et al., 2012). Studies by Zeng and Heenan have shown that the susceptibility of rice to salinity stress varies with growth stages (Heenan et al., 1988; Zeng et al., 2002). At the germination, active tillering and maturity rice is less sensitive to saline conditions compared to the vegetative and early seedling stages (Fageria, 1985; Heenan et al., 1988; Khan et al., 1997). While vegetative growth during the early seedling stage is highly sensitive to saline conditions, and often translates to reduced stand density and ultimately reduced yields in salt-affected fields (Lutts et al., 1995; Gregorio et al., 2002). Moreover, some rice varieties are most sensitive to salt stress during early tillering and panicle initiation stages of growth (Zeng et al., 2002).

1.1 Physiological responses to salinity

Considering the ubiquitous nature and detrimental effects of salinity on crop productivity, numerous studies have been conducted to characterize the physiological responses to salinity in an effort to improve tolerance to excessive salt. Salinity tolerance can be simply defined as the

ability to maintain growth or yield under saline conditions. However, the physiological responses to salinity, and the underlying tolerance mechanisms are much more complex. The reduction in growth in response to salinity is due to the osmotic effects of excessive ions in the soil, which reduce water uptake and cell expansion, and the ionic effects, which are largely the result of the toxic effects of excess sodium in the cells (Figure 1). These two phases are not mutually exclusive, but often occur at different times during salt stress, and involve very different tolerance mechanisms.

The predominance of the ionic or osmotic effects on plant growth will also depend on the duration of stress treatment and the concentration of NaCl. The osmotic effects of high Na^+ in the external media are likely to have a much greater impact on plant growth in low to moderate NaCl concentrations, compared to ionic effects (Munns and Tester, 2008). For growth to be reduced in the ionic phase Na^+ must accumulate to toxic levels to cause cell death and/or impede metabolic processes. This would likely be experienced only under high NaCl concentrations or long periods of salt stress. Here, the physiological responses and underlying tolerance mechanisms to each phase of the salt response (osmotic and ionic) are discussed.

1.1.1 Osmotic effects of salinity

The osmotic effects of excessive salt have almost an immediate impact on plant growth (Figure 1a) (Munns and Tester, 2008). The high concentration of soluble salts in the soil result in a reduction in soil water potential, which limits cellular water uptake and reduces cell expansion. In time, these will translate to reduced leaf growth, a reduction in the rate of lateral shoot development and/or growth, and a reduction in stomatal conductance and carbon sequestration.

Most plants will respond to a reduction in external water potential by lowering the cell water potential, typically through the production of solutes such as proline, trehalose, or the use of ions (Figure 2). This mechanism, known as osmotic adjustment, restores the gradient in water potential between the internal and external environments, and facilitates water uptake and cell expansion. Several studies have shown that genetic variation for osmotic adjustment during water deficits exist in rice germplasm (Lilley and Ludlow, 1996; Lilley et al., 1996; Babu et al., 1999). Lilley and Ludlow evaluated 61 diverse rice accessions under drought conditions and showed that *Indica* accessions displayed greater tolerance to dehydration stress and better osmotic adjustment compared to *Japonica* accessions (Lilley and Ludlow, 1996). Moreover, Lilley *et al* performed QTL mapping and identified single QTL that accounted for ~30% of phenotypic variation for osmotic adjustment during drought, suggesting that osmotic adjustment is heritable and may be improved using conventional breeding methods (Lilley et al., 1996). While osmotic adjustment has been shown to be a mechanism to tolerate drought conditions in rice, its importance in osmotic tolerance during of salt stress is currently limited.

1.1.2 Ionic effects of salinity

The ionic effects and the corresponding tolerance mechanisms have been well characterized in cereals. The toxic effect of sodium is largely due to its similar physiochemical properties with potassium. Potassium is an important cofactor for many cellular reactions, and serves as a major counter ion to balance negatively charged ions within the cell. Due to the similarities between sodium and potassium, sodium will complete with potassium binding sites in many enzymes and disrupt many cellular processes. Thus many studies have suggested that maintaining high K^+ levels relative to Na^+ may mitigate the toxic effects of sodium (Tester and Davenport, 2003).

Tolerance mechanisms to the ionic phase of salinity stress can be characterized into two classes: sodium exclusion and tissue tolerance (Munns and Tester, 2008) (Figure 2). Sodium exclusion mechanisms act to limit the accumulation of sodium in photosynthetic tissues, either through the storage of Na^+ in older leaves or through sequestration in the roots. Such mechanisms should preserve photosynthetically active tissues, thus maintaining carbon sequestration and plant growth. Numerous transporters have been identified that act in both the root and the shoot to effectively limit Na^+ accumulation, and are discussed in more detail below. Tissue tolerance refers to the ability to withstand high concentrations of Na^+ in the leaf tissue without leaf senescence. Here, Na^+ may be compartmentalized into the vacuole where it cannot interfere with the reactions occurring in the cytoplasm. However, compatible solutes (sucrose, proline) must be produced and accumulate in the cytoplasm to restore osmotic balance caused by an excess of Na^+.

1.2 Na^+ transport processes in roots

The high concentration of Na^+ in the soil solution, as well as the large voltage differences between the cell and external environment facilitate the passive uptake of Na^+. The total accumulation of Na^+ in the root is due to passive influx, as well as the active efflux of Na^+ from the cell. Most of the unidirectional influx of Na^+ into the cell is the result of transporters that move other cations such as $NH4^+$ or K^+, and are not selective to Na^+. Two types of transport processes mediate Na^+ influx across the plasmamembrane: Ca^{+2} sensitive and Ca^{2+} insensitive. The beneficial effects of Ca^{2+} on plant growth during salinity stress has been documented as early as 1902, however discoveries made in the last 20 years regarding the molecular mechanisms underlying salinity responses have shed additional light on Ca^{2+}-Na^+ interactions

(Cramer, 2002). It is likely that Ca^{2+} can mitigate the toxic effects of Na^{+} by inhibiting these Ca^{2+} sensitive influx mechanisms.

Non-selective cation channels (NSCCs), which include glutamate-activated channels (GLRs), cyclic nucleotide gated cation channels (CNGCs), and voltage-independent nonselective cation channels (VI-NSCC), are prime candidates for Ca^{2+}-sensitive Na^{+} uptake. In the case of VI-NSCCs, Na^{+} transport is directly blocked by Ca^{2+} (Demidchik and Maathuis, 2007). However, the protein families for NSCCs are often quite large in plant species, and therefore have hindered the study of individual gene members. For even the plant model Arabidopsis, only a small handful have been characterized and few have been shown to transport Na^{+}.

The *Salt Overly Sensitive (SOS)* pathway is involved in cellular Na^{+} homeostasis and is regulated by Ca^{+2}. While this pathway has been characterized in great detail in Arabidopsis, it is also important in the regulation of salinity responses in rice (Martinez-Atienza et al., 2007). The pathway is focused around a H^{+}/Na^{+} antiporter *SOS1*, which regulates the efflux of Na^{+} from the cytoplasm across the plasmamembrane (Luan, 2009). *SOS1* is regulated by a CBL-CIPK complex that is composed of a calcineurin B-like protein (*CBL4/SOS3*) and *SOS2* (*CIPK24*). After the addition of Na^{+} the initial burst of Ca^{2+} is percieved by a calcineurin B-like protein (*CBL4/SOS3*), which then recruits *SOS2* and phosphorylates *SOS1* (Luan, 2009). In addition to regulating *SOS1*, *SOS2* has been shown to control Na^{+} compartmentalization via the regulation of tonoplast H^{+}/Na^{+} antiporters and vacuolar *H^{+}-ATPases* (Qiu et al., 2004; Batelli et al., 2007).

The HKT gene family has been shown to play an important role in the regulation of Ca^{2+}-insensitive Na^{+} flux in many plant species (Platten et al., 2006; Hauser and Horie, 2010; Waters et al., 2013). In rice, the HKT gene family consists of eight members that can be subdivided into

two major classes based on Na^+ and K^+ selectivity (Platten et al., 2006). Class II members typically exhibit both Na^+ and K^+ transport (however there are exceptions see (Horie et al., 2009)), while members in class I typically exhibit Na^+ selective transport (Platten et al., 2006). Unlike, members of the SOS pathway and NSCCs several members of the HKT gene family have been identified as causal genes underlying salinity tolerance, and are discussed in further detail below (Lin et al., 2004; Ren et al., 2005; Huang et al., 2006; Rus et al., 2006; Munns et al., 2012).

The total Na^+ content in the root is dependent on the rate of passive Na influx from the soil solution into the root, and the rate of active Na efflux from the root into the soil solution. The efflux of Na^+ out of the root is most likely performed by a Na^+/H^+ antiporter, such as *SOS1*. Here, an electrochemical H^+ gradient is used to move Na^+ out of the cell. This gradient is generated by a H^+-ATPase, which utilizes ATP hydrolysis to pump H^+ out of the cell. While, Na^+-ATPases are important for Na^+ transport in animal systems, no Na^+-ATPases have been identified in plants, indicating that Na^+/H^+ system is the predominate mechanism in plants (Horie and Schroeder, 2004). However, it should be noted that once Na+ has been transported to the apoplast, passive unidirectional transport mechanisms would likely move Na+ back into the cell, thus creating a futile Na^+ cycle at the plasmamembrane. The importance of this mechanism in adaptation to saline conditions has been challenged, and it is questionable whether such secondary active transport mechanisms would be advantageous, from a metabolic perspective, under most salt stress conditions (Tester and Davenport, 2003; Kronzucker et al., 2006; Britto and Kronzucker, 2015).

1.3 Na^+ translocation within the plant

Limiting Na^+ concentration in the xylem is a central component of shoot sodium exclusion. In the root, Na^+ concentration in the xylem sap is effectively regulated by controlling the rates of xylem loading as well as the rates of xylem unloading. While few genes have been identified that regulate the loading of Na^+ into xylem vessels, it is thought that the Na^+/H^+ transport system described above is a likely candidate (Plette and Møller, 2010). However, numerous HKT gene members have been identified that regulate the retrieval of Na^+ from the xylem in root tissues (Ren et al., 2005; Byrt et al., 2007; Munns et al., 2012).

In rice, xylem unloading of Na^+ is largely achieved through the actions of *HKT1;5* (Ren et al., 2005). This gene was initially discovered using a biparental population generated from a cross between the a tolerant *Indica* accession 'Nona Bokra' and a sensitive *Japonica* accession 'Koshihikari' (Lin et al., 2004). *HKT1;5* is a bidirectional Na+ transporter that regulates both Na+ efflux and influx across the plasmamembrane in a K+-independent manner, and is predominately expressed in root tissue (Ren et al., 2005). The tolerant allele of HKT1;5 derived from 'Nona Bokra' display higher transport activity compared to the 'Koshihikari' allele, and are largely responsible for the lower shoot Na+ observed in 'Nona Bokra'.

A more recent study in wheat identified *TmHKT1;5*, which exhibits similar expression patterns and transport properties as *OsHKT1;5*, suggesting that these genes are likely orthologs (Munns et al., 2012). Interestingly, in this study tolerant alleles were derived from the diploid wheat relative *Triticum monococcum* and improved shoot Na^+ exclusion when introgressed into durum wheat varieties (*Triticum turgidum*). These studies highlight the importance of Na^+ retrieval from the xylem in roots for improving tolerance to the ionic phase of salt stress.

Once Na^+ has entered into the transpiration stream it can have several fates: it can be partitioned to within the shoot to minimize damage to photosynthetically active tissues, recirculated via the phloem, or can be compartmentalized within cells to minimize the damage leaf tissue (Munns and Tester, 2008). While direct evidence for each of these mechanisms is currently lacking in rice, studies in other cereals such as wheat or barley have provided evidence that these may be effective mechanisms to tolerate excessive Na^+.

In wheat, a study by Lindsay *et al* identified *Nax1* as an important locus in the regulation of shoot Na^+ exclusion in durum wheat (Lindsay et al., 2004). This QTL accounted for ~38% of the phenotypic variation for Na^+ concentration in the leaf blade in the mapping population (Lindsay et al., 2004). In a subsequent study, James *et al* further characterized near isogenic lines and found that lines that harbored *Nax1* retained a high sheath to blade concentration of Na^+ (James et al., 2006). This indicated that *Nax1* may unload Na^+ from the xylem stream in leaf sheaths, in an effort to reduce the amount of Na^+ that is accumulated in active photosynthetic leaf blades. Fine mapping and expression profiling identified *TmHKT7* as a possible candidate for *Nax1* (Huang et al., 2006). The expression patterns of *TmHKT7* matched the putative physiological role suggested by James *et al*.

Although Na^+ recirculation within the plant likely plays a role in the shoot Na exclusion, the importance and the underlying mechanisms have remained a highly debated topic among researchers. Several studies have shown that in many species Na is recirculated from the shoot to the roots via phloem (see Tester and Davenport, 2003; Plette and Møller, 2010 for discussion), however limiting the entry of Na^+ into the xylem is likely a much more important and effective mechanism for limiting shoot Na^+ content. In rice, Wang et al characterized a T-DNA knockout

mutant for HKT1;1 and found that mutants accumulate Na^+ in xylem sap and display a reduction in Na^+ in phloem sap compared to wild-type. These observations prompted Wang *et al* to suggest that *HKT1;1* may regulate sodium exclusion from the shoot of seedlings possibly through xylem-to-phloem or parenchyma-to-xylem transfer of Na^+ (Wang et al., 2015). However, further experimental evidence is needed to substantiate this model.

In Arabidopsis, the model of shoot-to-root recirculation has been a topic of fierce discussion and the contrasting results of independent studies have left a complicated and confusing story. Like *OsHKT1;1* knockout mutants, in Arabidopsis *athkt1;1* knockout mutants also exhibit higher xylem Na^+ and lower phloem Na^+ (Berthomieu et al., 2003; Sunarpi et al., 2005; Wang et al., 2015), and it was initially proposed that was*AtHKT1;1* functioned in the recirculation of Na^+ from the root to the shoot (via loading of Na^+ into the phloem in the shoots). However, work by Sunarpi *et al* later proposed that *AtHKT1;1* functions primarily in the removal of Na^+ from the xylem sap and eventually to the phloem largely through symplastic diffusion (Berthomieu et al., 2003; Sunarpi et al., 2005). A later study presented very different results and showed that *AtHKT1;1* was primarily involved in the retrieval of Na^+ from the xylem in root tissue, and suggested that the function of AtHKT1;1 in shoot tissue may be dependent on the experimental conditions (discussed in (Munns and Tester, 2008)) (Davenport et al., 2007). Therefore, although it is plausible that shoot-to-root recirculation may play a role in sodium exclusion in the shoot, the importance and underlying mechanisms require further study.

1.4 Current challenges and future directions for improving salt tolerance

1.4.1 Assessing the temporal nature of salt response

Salinity tolerance is a genetically complex trait (Roy et al., 2014). The maintenance of growth and yield under saline conditions is a combined result of many biological pathways and numerous genes (Horie and Schroeder, 2004; Munns and Tester, 2008; Horie et al., 2009; Roy et al., 2014). This is not surprising considering the breadth of physiological mechanisms plants use to adapt to salinity. While conventional breeding approaches that have selected for yield or biomass under saline conditions have proven to be largely ineffective for developing salt tolerant varieties, programs that have focused on individual adaptive traits, such as sodium exclusion, have identified numerous genes that can be introgressed into commercial varieties (Ren et al., 2005; Byrt et al., 2007; Munns et al., 2012). However, these successes have largely focused on the ionic component of the salt response. Currently, no studies have identified genes that have a major contribution to tolerance to the osmotic phase of the salt response (Roy et al., 2014).

The temporal nature of the physiological responses to salinity remains a major challenge in identifying the genetic basis of tolerance to the ionic and osmotic effects of excess salt. To assess tolerance to each individual phase, the researcher must know when to destructively sample for ion or biomass traits. This is particularly problematic given the size and diversity of many mapping populations. While it is advantageous to increase genetic diversity within the mapping population, the high diversity will likely also mean that the threshold to ion toxicity may vary considerably between accessions. Therefore, after a week of salinity stress some accessions may be experiencing the osmotic effects of salinity, while others may be experiencing the ionic

effects of salinity. Moreover, the large size of most mapping populations necessitates considerably resources (space, labor) for multiple destructive measurements.

The advent of high-throughput image-based phenotyping platforms helps overcome many of these challenges. With these systems, researchers can nondestructively quantify plant growth, chlorophyll health/leaf senescence, canopy/leaf temperature, and mineral nutrition (Berger et al., 2010; Golzarian et al., 2011; Hairmansis et al., 2014; Honsdorf et al., 2014; Campbell et al., 2015, 2017; Al-Tamimi et al., 2016). The multiple imaging systems, as well as the introduction of the time axis allows multiple physiological responses to be assessed throughout the duration of the experiment. Moreover, combinations of imaging systems allow the ion-dependent and independent effects of salinity to be quantified. These methods facilitate the development of functional models that describe the responses to saline conditions, and may provide new insights into the genetic mechanisms of salt tolerance provided by studies using conventional phenotyping.

1.4.2 Bridging the genotype-to-phenotype gap

Over the past decade genome-wide association studies (GWAS) have proven to be a powerful approach to identify genetic variants associated with agronomically important traits (Huang et al., 2010; Zhang et al., 2010; Morris et al., 2013; Yang et al., 2014; Crowell et al., 2016; McCouch et al., 2016). Recent studies in rice have utilized GWAS to examine the genetic architecture of relative growth rate, transpiration rate, transpiration use efficiency (TUE), and ion content during salinity stress (Campbell et al., 2015; Al-Tamimi et al., 2016; Patishtan et al., 2017). While these studies have identified numerous, potentially agrononically important loci,

the identification of causal genes and functional DNA elements underlying these traits remains a major challenge in uncovering the genetic mechanisms that contribute to phenotypic variation.

Following the central dogma of molecular biology, causal variants should change the production of RNA and/or protein and lead to perturbation in the phenotype. Thus, the utilization of intermediate phenotypes, such as gene expression, metabolites, or protein levels, is a natural step forward in identifying the causal variants underlying phenotypic variation (Pasaniuc and Price, 2016). RNA expression can be quantified from a diverse population of individuals, and genetic variants that may affect gene expression (eQTL) can be identified using the standard statistical framework for association mapping (Ingvarsson and Street, 2011). The colocalization of eQTL and phenotypic QTL (pQTL) can greatly accelerate the selection of potential causal variants and/or genes that can be functionally validated through transgenic approaches (Giambartolomei et al., 2014; Gusev et al., 2016; Pasaniuc and Price, 2016; Zhu et al., 2016).

While it seems straightforward to identify colocalized GWAS signals by examining overlapping QTLs, these common associations may by due largely to LD rather than a common causal variant (Giambartolomei et al., 2014; Gusev et al., 2016; Zhu et al., 2016). Several approaches have been developed to integrate signals from GWAS and eQTL studies and distinguish between regulation of the trait and gene expression by common causal variant and LD (Giambartolomei et al., 2014; Gusev et al., 2016; Zhu et al., 2016). These approaches have been successfully implemented in human GWAS studies to identify genes potentially underlying loci associated with diseases (Onengut-Gumuscu et al., 2015; Pavlides et al., 2016; Civelek et al., 2017; Mancuso et al., 2017).

The TWAS method developed by Gusev *et al*, utilizes expression measurements collected from a relatively small sample of genotyped individuals, and imputes the *cis*-genetic component of expression on a much larger set of individuals for which phenotype and genotype data exists (Gusev et al., 2016). The approach correlates the imputed gene expression levels with the trait, yielding gene-trait associations. The particular advantage of this method is that for the large phenotyped population, only summary statistics from a GWAS are necessary.

In addition to TWAS, which focuses largely on gene expression data, the methodology developed by Giambartolomei *et al* utilizes summary GWAS statistics from any two traits to determine if they are regulated by a common variant (Giambartolomei et al., 2014). This method takes a Bayesian approach to determine the posterior probabilities of five hypothesis for each QTL and trait pair: 1) No association with either trait; 2) the SNP is associated with trait 1, but not trait 2; 3) the SNP is associated with trait 2, but not trait 1; 4) there is an association within the region for both traits 1 and 2, but they regulated by two independent causal variants; 5) both traits are regulated by a single common causal variant. Similar to the TWAS approach, this method can be run using only summary statistics from GWAS studies, and it accounts for local LD at the region of interest.

While these methods have proven to be useful strategies for dissecting the genetic basis of complex traits in humans, mice and yeast, the lack of large-scale expression data for rice and other crops have hindered the use of these techniques for studying the genetics of agronomically important traits. However, as the price of next generation sequencing continues to decline and the availability of more cost-effective methods to assess gene expression these approaches can be directly applied to crops to identify genes for functional validation.

1.5 Aims of the current study

Despite intensive investigations of genetic variation in salinity tolerance in rice, few genes have been identified that regulate salinity tolerance-related traits that can be used by breeders to introgress tolerant alleles into commercial cultivars. Given the diverse adaptive strategies to salinity and the complex genetic basis of salt tolerance, a genetical genomics approach that integrates GWAS signals derived from various phenotypes and eQTL loci will facilitate the identification of new loci that underlie salt tolerance-related traits.

In this study, the genetic basis of salt tolerance in rice was explored using a multi-disciplinary approach that combines aspects of plant physiology, molecular biology, and quantitative genetics, and computer science. The overall aim of these studies is to dissect the genetic basis of salt tolerance in rice in an effort to identify candidate genes that underlie genetic variation for salt tolerance-related traits. To this end, a publically available germplasm collection was utilized. This panel contains ~406 diverse rice accessions, and captures much of the genetic variation in cultivated rice (Zhao et al., 2011). The specific aims of this dissertation are: (1) to assess salt tolerance using conventional and high-throughput phenotyping approaches; (2) develop novel, high-throughput methods that can extract biologically-relevant traits from images derived from image-based phenomic platforms; (3) develop statistical methods to functionally model longitudinal growth and growth-response, and identify loci associated with variation in trajectories; (4) identify pathways that are affected by short term salinity stress in shoot tissue; and (5) utilize large-scale transcriptome data to identify *cis* regulatory variants and determine their contribution to salt tolerance-related traits.

References

Al-Tamimi, N., C. Brien, H. Oakey, B. Berger, S. Saade, Y.S. Ho, S.M. Schmöckel, M. Tester, and S. Negrão. 2016. Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. Nat. Commun. 7.

Babu, R.C., M.S. Pathan, A. Blum, and H.T. Nguyen. 1999. Comparison of measurement methods of osmotic adjustment in rice cultivars. Crop Sci. 39(1): 150–158.

Batelli, G., P.E. Verslues, F. Agius, Q. Qiu, H. Fujii, S. Pan, K.S. Schumaker, S. Grillo, and J.-K. Zhu. 2007. SOS2 promotes salt tolerance in part by interacting with the vacuolar H+-ATPase and upregulating its transport activity. Mol. Cell. Biol. 27(22): 7781–7790.

Berger, B., B. Parent, and M. Tester. 2010. High-throughput shoot imaging to study drought responses. J. Exp. Bot. 61(13): 3519–28.

Berthomieu, P., G. Conéjéro, A. Nublat, W.J. Brackenbury, C. Lambert, C. Savio, N. Uozumi, S. Oiki, K. Yamada, F. Cellier, F. Gosti, T. Simonneau, P. a. Essah, M. Tester, A.A. Véry, H. Sentenac, and F. Casse. 2003. Functional analysis of AtHKT1 in Arabidopsis shows that Na+ recirculation by the phloem is crucial for salt tolerance. EMBO J. 22(9): 2004–2014.

Britto, D.T., and H.J. Kronzucker. 2015. Sodium efflux in plant roots: What do we really know? J. Plant Physiol. 186: 1–12.

Byrt, C.S., J.D. Platten, W. Spielmeyer, R.A. James, E.S. Lagudah, E.S. Dennis, M. Tester, and R. Munns. 2007. HKT1; 5-like cation transporters linked to Na+ exclusion loci in wheat, Nax2 and Kna1. Plant Physiol. 143(4): 1918–1928.

Campbell, M.T., Q. Du, K. Liu, C. Brien, B. Berger, C. Zhang, and H. Walia. 2017. A Comprehensive Image-Based Phenomics Analysis Reveals The Complex Genetic Architecture Of Shoot Growth Dynamics In Rice (Oryza Sativa). Plant Genome IN PRESS.

Campbell, M.T., A.C. Knecht, B. Berger, C.J. Brien, D. Wang, and H. Walia. 2015. Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. Plant Physiol. 168(August): pp.00450.2015.

Chinnusamy, V., a Jagendorf, and J.K. Zhu. 2005. Understanding and improving salt tolerance in plants. Crop Sci. 45: 437–448.

Civelek, M., Y. Wu, C. Pan, C.K. Raulerson, A. Ko, A. He, C. Tilford, N.K. Saleem, A. Stančáková, L.J. Scott, and others. 2017. Genetic regulation of adipose gene expression and cardio-metabolic traits. Am. J. Hum. Genet. 100(3): 428–443.

Cramer, G.R. 2002. Sodium-calcium interactions under salinity stress. p. 205–227. *In* Salinity: Environment-Plants-Molecules. Springer.

Crowell, S., P. Korniliev, A. Falcão, A. Ismail, G. Gregorio, J. Mezey, and S. McCouch. 2016. Genome-wide association and high-resolution phenotyping link Oryza sativa panicle traits to numerous trait-specific QTL clusters. Nat. Commun. 7.

Davenport, R.J., A. Muñoz-Mayor, D. Jha, P.A. Essah, A. Rus, and M. Tester. 2007. The Na+

transporter AtHKT1;1 controls retrieval of Na + from the xylem in Arabidopsis. Plant, Cell Environ. 30(2007): 497–507.

Demidchik, V., and F.J.M. Maathuis. 2007. Physiological roles of nonselective cation channels in plants: from salt stress to signalling and development. New Phytol. 175(3): 387–404.

Fageria, N.K. 1985. Salt tolerance of rice cultivars. Plant Soil 88(2): 237–243.

Flowers, T.J., and A.R. Yeo. 1995. Breeding for salinity resistance in crop plants: where next? Funct. Plant Biol. 22(6): 875–884.

Giambartolomei, C., D. Vukcevic, E.E. Schadt, L. Franke, A.D. Hingorani, C. Wallace, and V. Plagnol. 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10(5): e1004383.

Golzarian, M.R., R. a Frick, K. Rajendran, B. Berger, S. Roy, M. Tester, and D.S. Lun. 2011. Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods 7(1): 2.

Gregorio, G.B., D. Senadhira, R.D. Mendoza, N.L. Manigbas, J.P. Roxas, and C.Q. Guerta. 2002. Progress in breeding for salinity tolerance and associated abiotic stresses in rice. F. Crop. Res. 76(2): 91–101.

Gusev, A., A. Ko, H. Shi, G. Bhatia, W. Chung, B.W.J.H. Penninx, R. Jansen, E.J.C. De Geus, D.I. Boomsma, F.A. Wright, and others. 2016. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet.

Hairmansis, A., B. Berger, M. Tester, and S.J. Roy. 2014. Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice. Rice 7(1): 16.

Hauser, F., and T. Horie. 2010. A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K+/Na+ ratio in leaves during salinity stress. Plant. Cell Environ. 33(4): 552–565.

Heenan, D.P., L.G. Lewin, and D.W. McCaffery. 1988. Salinity tolerance in rice varieties at different growth stages. Aust. J. Exp. Agric. 28(3): 343–349.

Honsdorf, N., T. March, B. Berger, M. Tester, and K. Pillen. 2014. High-Throughput Phenotypig to Detect Drought Tolerance QTL in Barley Introgression Lines. PLoS One 9(5): e97047.

Horie, T., F. Hauser, and J.I. Schroeder. 2009. HKT transporter-mediated salinity resistance mechanisms in Arabidopsis and monocot crop plants. Trends Plant Sci. 14(12): 660–668.

Horie, T., and J.I. Schroeder. 2004. Sodium transporters in plants. Diverse genes and physiological functions. Plant Physiol. 136(1): 2457–2462.

Huang, S., W. Spielmeyer, E.S. Lagudah, R.A. James, J.D. Platten, E.S. Dennis, and R. Munns. 2006. A sodium transporter (HKT7) is a candidate for Nax1, a gene for salt tolerance in durum wheat. Plant Physiol. 142(4): 1718–1727.

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, M. Li, D. Fan, Y. Guo, A. Wang, L. Wang, L. Deng, W. Li, Y. Lu, Q. Weng, K. Liu, T. Huang, T.

Zhou, Y. Jing, W. Li, Z. Lin, E.S. Buckler, Q. Qian, Q. Zhang, J. Li, and B. Han. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Publ. Gr. 42(11): 961–967Available at http://dx.doi.org/10.1038/ng.695.

Ingvarsson, P.K., and N.R. Street. 2011. Association genetics of complex traits in plants. New Phytol. 189(4): 909–922.

James, R.A., R.J. Davenport, and R. Munns. 2006. Physiological characterization of two genes for Na+ exclusion in durum wheat, Nax1 and Nax2. Plant Physiol. 142(4): 1537–1547.

Khan, M.S.A., A. Hamid, and M.A. Karim. 1997. Effect of sodium chloride on germination and seedling characters of different types of rice (Oryza sativa L.). J. Agron. Crop Sci. 179(3): 163–169.

Kronzucker, H.J., M.W. Szczerba, M. MOAZAMI-GOUDARZI, and D.T. Britto. 2006. The cytosolic Na+: K+ ratio does not explain salinity-induced growth impairment in barley: a dual-tracer study using 42K+ and 24Na+. Plant. Cell Environ. 29(12): 2228–2237.

Lilley, J.M., and M.M. Ludlow. 1996. Expression of osmotic adjustment and dehydration tolerance in diverse rice lines. F. Crop. Res. 48(2): 185–197.

Lilley, J.M., M.M. Ludlow, S.R. McCouch, and J.C. O'toole. 1996. Locating QTL for osmotic adjustment and dehydration tolerance in rice. J. Exp. Bot. 47(9): 1427–1436.

Lin, H.X., M.Z. Zhu, M. Yano, J.P. Gao, Z.W. Liang, W. a Su, X.H. Hu, Z.H. Ren, and D.Y. Chao. 2004. QTLs for Na+ and K+ uptake of the shoots and roots controlling rice salt tolerance. Theor. Appl. Genet. 108(2): 253–60.

Lindsay, M.P., E.S. Lagudah, R.A. Hare, and R. Munns. 2004. A locus for sodium exclusion (Nax1), a trait for salt tolerance, mapped in durum wheat. Funct. Plant Biol. 31(11): 1105–1114.

Luan, S. 2009. The CBL--CIPK network in plant calcium signaling. Trends Plant Sci. 14(1): 37–42.

Lutts, S., J.M. Kinet, and J. Bouharmont. 1995. Changes in plant response to NaCl during development of rice (Oryza sativa L.) varieties differing in salinity resistance. J. Exp. Bot. 46(12): 1843–1852.

Mancuso, N., H. Shi, P. Goddard, G. Kichaev, A. Gusev, and B. Pasaniuc. 2017. Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. Am. J. Hum. Genet. 100(3): 473–487.

Martinez-Atienza, J., X. Jiang, B. Garciadeblas, I. Mendoza, J.-K. Zhu, J.M. Pardo, and F.J. Quintero. 2007. Conservation of the salt overly sensitive pathway in rice. Plant Physiol. 143(2): 1001–1012.

McCouch, S.R., M.H. Wright, C.-W. Tung, L.G. Maron, K.L. McNally, M. Fitzgerald, N. Singh, G. DeClerck, F. Agosto-Perez, P. Korniliev, and others. 2016. Open access resources for genome-wide association mapping in rice. Nat. Commun. 7.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, and S. Kresovich. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc. Natl. Acad. Sci. 110(2): 453–8Available at http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3545811&tool=pmcentrez&rendertype=abstract.

Munns, R., R. a James, B. Xu, A. Athman, S.J. Conn, C. Jordans, C.S. Byrt, R. a Hare, S.D. Tyerman, M. Tester, D. Plett, and M. Gilliham. 2012. Wheat grain yield on saline soils is improved by an ancestral Na^+ transporter gene. Nat. Biotechnol. 30(4): 360–4.

Munns, R., and M. Tester. 2008. Mechanisms of salinity tolerance. Annu. Rev. Plant Biol. 59: 651–81.

Onengut-Gumuscu, S., W.-M. Chen, O. Burren, N.J. Cooper, A.R. Quinlan, J.C. Mychaleckyj, E. Farber, J.K. Bonnie, M. Szpak, E. Schofield, and others. 2015. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat. Genet. 47(4): 381–386.

Pasaniuc, B., and A.L. Price. 2016. Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet.

Patishtan, J., T.N. Hartley, R.F. de Carvalho, and F.J.M. Maathuis. 2017. Genome Wide Association Studies To Identify Rice Salt-Tolerance Markers. Plant. Cell Environ.Available at http:https://dx.doi.org/10.1111/pce.12975.

Pavlides, J.M.W., Z. Zhu, J. Gratten, A.F. McRae, N.R. Wray, and J. Yang. 2016. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. Genome Med. 8(1): 84.

Platten, J.D., O. Cotsaftis, P. Berthomieu, H. Bohnert, R.J. Davenport, D.J. Fairbairn, T. Horie, R.A. Leigh, H.-X. Lin, S. Luan, and others. 2006. Nomenclature for HKT transporters, key determinants of plant salinity tolerance. Trends Plant Sci. 11(8): 372–374.

Plette, C.D., and I.S. Møller. 2010. Na+ transport in glycophytic plants: what we know and would like to know. Plant. Cell Environ. 33(4): 612–626.

Qiu, Q.-S., Y. Guo, F.J. Quintero, J.M. Pardo, K.S. Schumaker, and J.-K. Zhu. 2004. Regulation of vacuolar Na+/H+ exchange in Arabidopsis thaliana by the salt-overly-sensitive (SOS) pathway. J. Biol. Chem. 279(1): 207–215.

Redfern, S.K., N. Azzu, and J.S. Binamira. 2012. Rice in Southeast Asia: facing risks and vulnerabilities to respond to climate change. Build. Resil. Adapt. to Clim. Chang. Agric. Sect. 23: 295.

Ren, Z.-H., J.-P. Gao, L.-G. Li, X.-L. Cai, W. Huang, D.-Y. Chao, M.-Z. Zhu, Z.-Y. Wang, S. Luan, and H.-X. Lin. 2005. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat. Genet. 37(10): 1141–6.

Roy, S.J., S. Negrão, and M. Tester. 2014. Salt resistant crop plants. Curr. Opin. Biotechnol. 26:

115–124.

Rus, A., I. Baxter, B. Muthukumar, J. Gustin, B. Lahner, E. Yakubova, and D.E. Salt. 2006. Natural variants of AtHKT1 enhance Na+ accumulation in two wild populations of Arabidopsis. PLoS Genet. 2(12): 1964–1973.

Sunarpi, T. Horie, J. Motoda, M. Kubo, H. Yang, K. Yoda, R. Horie, W.-Y. Chan, H.-Y. Leung, K. Hattori, M. Konomi, M. Osumi, M. Yamagami, J.I. Schroeder, and N. Uozumi. 2005. Enhanced salt tolerance mediated by AtHKT1 transporter-induced Na unloading from xylem vessels to xylem parenchyma cells. Plant J. 44(6): 928–38.

Tester, M., and R. Davenport. 2003. Na+ tolerance and Na+ transport in higher plants. Ann. Bot. 91(5): 503–527.

Wang, R., W. Jing, L. Xiao, Y. Jin, L. Shen, and W. Zhang. 2015. The rice high-affinity potassium Transporter1; 1 is involved in salt tolerance and regulated by an MYB-type transcription factor. Plant Physiol. 168(3): 1076–1090.

Waters, S., M. Gilliham, and M. Hrmova. 2013. Plant high-affinity potassium (HKT) transporters involved in salinity tolerance: structural insights to probe differences in ion selectivity. Int. J. Mol. Sci. 14(4): 7660–7680.

Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang, W. Fang, H. Feng, W. Xie, X. Lian, G. Wang, Q. Luo, Q. Zhang, Q. Liu, and L. Xiong. 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat. Commun. 5: 5087.

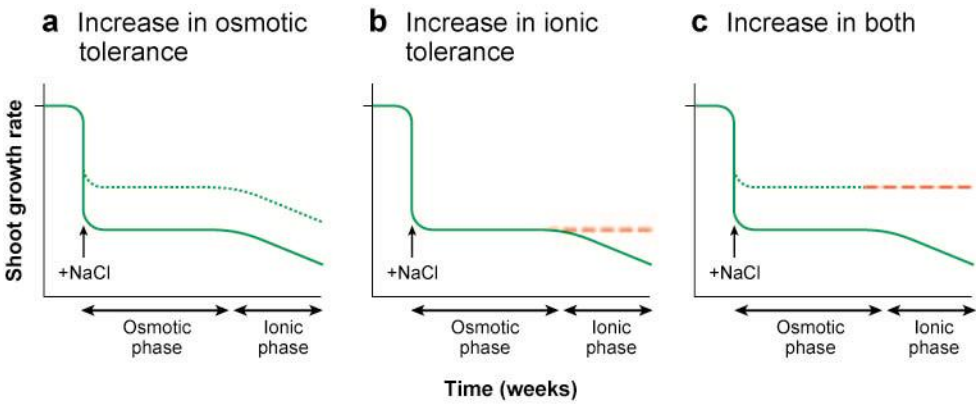
Zeng, L., M.C. Shannon, and C.M. Grieve. 2002. Evaluation of salt tolerance in rice genotypes by multiple agronomic parameters. : 235–245.

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and others. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42(4): 355–360.

Zhao, K., C.-W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2(1–4): 467.

Zhu, Z., F. Zhang, H. Hu, A. Bakshi, M.R. Robinson, J.E. Powell, G.W. Montgomery, M.E. Goddard, N.R. Wray, P.M. Visscher, and others. 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet.

1.6 Figures



Munns R, Tester M. 2008.
Annu. Rev. Plant Biol. 59:651–81.

Figure 1. Hypothetical shoot growth responses to salinity stress. The broken lines shows plant having a tolerant response to the osmotic phase of salt stress (a), the ionic phase of salt stress (b), or (c) showing tolerance to both phases of salt stress.

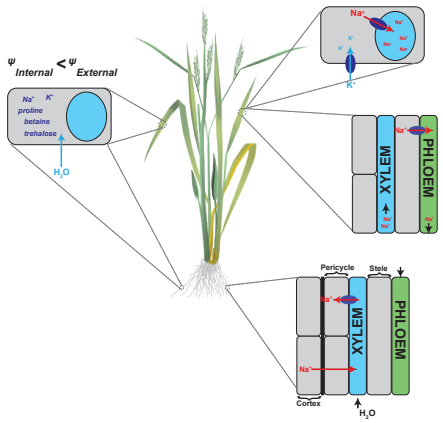


Figure 2. Schematic illustrating the major tolerance mechanisms documented in cereals. Here, not all mechanisms have been confirmed in rice, however these mechanisms have been widely studied in other cereals. Ψ refers to water potential.

CHAPTER 2

ALLELIC VARIANTS OF *OSHKT1;1* UNDERLIE THE DIVERGENCE BETWEEN *INDICA* AND *JAPONICA* SUBSPECIES OF RICE (*ORYZA SATIVA*) FOR ROOT SODIUM CONTENT

Malachy T Campbell, Nonoy Bandillo, Fouad Razzaq A Al Shiblawi, Sandeep Sharma, Kan Liu, Qian Du, Aaron J Schmitz, Chi Zhang, Anne-Aliénor Véry, Aaron J Lorenz, Harkamal Walia

Published: June 5, 2017; https://doi.org/10.1371/journal.pgen.1006823

Abstract

Salinity is a major limiting factor for crop productivity. Rice (*Oryza sativa*), a staple crop for the majority of the world, is highly sensitive to salinity stress. To discover novel sources of genetic variation for salt tolerance-related traits in rice, we screened 390 diverse accessions under 14 days of moderate (9 $dS \cdot m^{-1}$) salt stress. Considerable natural variation was observed among the major subpopulations for root Na^+ content, with *indica* accessions generally displaying higher root Na^+ and *japonica* accessions exhibiting lower root Na^+ content. The genetic basis of the observed variation in salinity phenotypes was elucidated through genome-wide association (GWA) using 397,812 single nucleotide polymorphisms. The strongest associations were identified for root Na^+:K^+ ratio and root Na^+ content in a region spanning ~575 Kb on chromosome 4, named *Root Na^+ Content 4* (*RNC4*). Two Na^+ transporters, *HKT1;1* and *HKT1;4* were identified as candidates for *RNC4.* Reduced expression of both *HKT1;1* and *HKT1;4* through RNA interference indicated that *HKT1;1* regulates shoot and root Na^+ content, and is likely the causal gene underlying *RNC4*. Three non-synonymous mutations within

HKT1;1 were present at higher frequency in the *indica* subpopulation. When expressed in *Xenopus* oocytes the predominant *indica* isoform exhibited higher inward (negative) currents and a less negative voltage threshold of inward rectifying current activation compared to the *japonica* predominant isoform. The introduction of a 4.5kb fragment containing the HKT1;1 promoter and CDS from an *indica* variety into a *japonica* background, resulted in a phenotype similar to the *indica* subpopulation, with higher root Na^+ and Na^+:K^+. Shoot growth responses to moderate levels of salinity were independent of tissue ion content. This study provides evidence that *HKT1;1* regulates root Na^+ content, and underlies the divergence in root Na^+ content between the two major subspecies in rice.

2.1 Introduction

Salinity is a widespread limitation for agricultural productivity, especially for irrigated agriculture and coastal lowlands prone to seawater ingress (Ismail et al., 2008; Ismail and Tuong, 2009). By definition, salinity occurs when there is a high concentration of soluble salts in soil (Munns and Tester, 2008). More than 800 million hectares worldwide is affected by salt, which accounts for 6% of the total land area (Munns and Tester, 2008). Besides natural causes such as rising sea levels during the dry and wet cropping seasons, the poor quality of irrigation water and improper drainage, also collectively increases soluble salt concentration in the root zone (Ismail and Tuong, 2009; Ismail et al., 2010).

Rice (*Oryza sativa* L.) is one of the most important crop species and is a staple food for more than half of the world's population. Salinity is a major impediment to increasing production in many rice growing regions, including temperate and tropical environments, around the world (Flowers and Yeo, 1995; Chinnusamy et al., 2005). Rice is the most salt-sensitive

species among major cereal crops (Munns and Tester, 2008). The susceptibility of rice to salinity stress varies with growth stages (Heenan et al., 1988; Zeng et al., 2002). Rice is less sensitive to saline conditions at germination, active tillering and maturity stage (Fageria, 1985; Heenan et al., 1988; Khan et al., 1997). Vegetative growth during the early seedling stage is highly sensitive to saline conditions, and often translates to reduced stand density in salt-affected fields (Lutts et al., 1995; Gregorio et al., 2002). Some rice varieties are most sensitive to salt stress during early tillering and panicle initiation stages of growth (Zeng et al., 2002). This developmentally-dependent salt-sensitivity, in context of yield reduction, was associated with significant decrease in tiller number per plant, spikelet number per panicle, fertility, panicle length and primary branches per panicle (Heenan et al., 1988; Cui et al., 1995; Khatun and Flowers, 1995; Zeng et al., 2002).

Despite the overall high salt-sensitivity of rice, several studies have demonstrated that considerable natural variation for salinity tolerance exists in rice germplasm (Senadhira, 1994; Negrão et al., 2011). Traditional landraces or cultivars such as ‘Pokkali’, ‘Nona Bokra’, ‘Cheriviruppu’ and ‘SR26B’ have originated or have been selected in coastal regions and are more tolerant to saline conditions (Flowers and Yeo, 1995; Gregorio et al., 2002; Flowers, 2004; Thomson et al., 2010). Quantitative trait loci (QTL) underlying salinity tolerance have undergone intensive investigations (Koyama et al., 2001; Bonilla et al., 2002; Lin et al., 2004; Thomson et al., 2010; Ahmadi et al., 2011; Negrão et al., 2011; Bandillo et al., 2013; Platten et al., 2013). Although many QTL have been identified across the rice genome, the most well-characterized QTL is *Saltol/SKC1*, which harbors *HKT1;5*, on the short arm of chromosome 1 (Bonilla et al., 2002; Lin et al., 2004; Thomson et al., 2010; Ahmadi et al., 2011). The *SKC1* gene (*HKT1;5*) was subsequently cloned from a salt-tolerant *indica* landrace, ‘Nona Bokra’, and

encodes a Na^+ transporter that regulates shoot Na^+:K^+ homeostasis during salt stress (Ren et al., 2005).

Salinity tolerance is a complex polygenic trait, and several physiological mechanisms, including tissue tolerance, sodium exclusion, osmotic stress tolerance, and tissue-specific sodium sequestration can be utilized for improving salinity tolerance (Munns and Tester, 2008). While many genes and QTL have been reported for salinity tolerance in rice, few studies have confirmed the importance of these resources for improving salinity tolerance. Hence, the genetic resources (QTL and genes) available to rice breeders for improving salt tolerance are limited. Identification of loci that regulate salt accumulation and/or distribution will enable the introgression of favorable genic combinations and greatly accelerate the development of robust salt-tolerant rice varieties.

Genetic variation within the rice germplasm collection can be utilized to identify important loci controlling variation for salinity tolerance through genome-wide association (GWA) analysis, which provides greater mapping resolution and evaluates greater allelic diversity compared to linkage mapping strategies (Yu and Buckler, 2006; Myles et al., 2009; Negrão et al., 2011; Platten et al., 2013). In this study, we used GWA to investigate the genetic architecture of salinity tolerance using the Rice Diversity Panel 1 (RDP1) (Zhao et al., 2011; Famoso et al., 2011; Eizenga et al., 2014). RDP1 is comprised of 421 accessions collected from 85 countries and was developed to identify alleles associated with morphological, physiological and agronomic traits (Zhao et al., 2011; Famoso et al., 2011; Eizenga et al., 2014). RDP1 captures much of the diversity in the rice germplasm collection worldwide (Zhao et al., 2011; Famoso et al., 2011; Eizenga et al., 2014). We used several quantitative measures to characterize

the rice diversity panel for physiological and morphological responses to salinity stress. Here, we show that allelic variants of a sodium transporter (*HKT1;1*) underlie natural variation for root Na^+ content in rice. Using transgenic plants, we demonstrate that variants within HKT1;1 alter Na^+ transport and can explain the basis of divergence in root Na^+ content between the *indica* and *japonica* subspecies of cultivated rice.

2.2 Results

To assess the degree of natural variation for salinity tolerance associated traits in rice, a 9 $dS \cdot m^{-1}$ (~90 mM NaCl) salt stress was imposed gradually over a period of four days (in four increments of 20-30 mM) to two-week old rice seedlings. A rice diversity panel, consisting of 390 rice accessions (383 from RDP1 and seven check varieties), was scored for ten phenotypic traits at the end of a two-week 9 $dS \cdot m^{-1}$ stress period (the plants were 28 days old). The ten traits recorded were root biomass (control and salt conditions), shoot biomass (control and salt conditions), root and shoot Na^+ content, root and shoot K^+ content, and root and shoot Na^+:K^+ (Supplemental File S1). To control for inherent differences in growth rate between lines, we expressed the saline-induced growth response for each accession as the ratio of biomass in saline conditions over biomass in control conditions. Broad sense heritability for the eight phenotypic traits ranged from 0.32 for root K^+ content and 0.83 for shoot biomass in control conditions (Supplemental Table S1).

2.2.1 Rice subpopulations exhibit inherent differences in ion homeostasis

To examine the relationships between each of the eight traits, Pearson correlation analysis was performed across all accessions. No significant relationship was observed between shoot biomass and ion traits (Supplemental Table S3). Moreover, root and shoot ion content

showed no significant relationship when the analysis was performed with all accessions (Supplemental Table S3). Due to the deep population structure in rice, correlation analysis was also performed for each of the five major subpopulations in RDP1 (here, admixed accessions were considered a separate subpopulation; Supplemental Table S4-S8) (Zhao et al., 2010; Famoso et al., 2011). Root growth response (the ratio of root biomass in salt to control) showed a weak negative correlation with shoot Na^+:K^+ in *admix* and *tropical japonica* (*trj*) accessions (Supplemental Tables S4 and S5, respectively). In *trj*, *aus*, and *tej* subpopulations significant, albeit weak, positive correlations were observed between shoot Na^+ and root Na^+:K^+ (Supplemental Tables S5-7).

Comparisons between each of the subpopulations showed significant differences for shoot and root Na^+, K^+ and Na^+:K^+ (Fig 1). *Indica* accessions exhibited significantly higher root Na^+ content and Na^+:K^+ compared to the other four subpopulations (Fig 1 A-B). Significantly lower shoot Na^+ and Na^+:K^+ were observed in *indica* and *aus* subpopulations compared to *temperate japonica* (*tej*), *tropical japonica* (*trj*) and *admix* accessions (Fig 1 C-D). These results suggest that there are inherent differences in root and shoot ion homeostasis between subpopulations, with *indica* accessions generally displaying higher root Na^+ and Na^+:K^+, and *indica* and *aus* accessions exhibiting lower shoot Na^+ content and Na^+:K^+.

2.2.2 Genome-wide association mapping identifies a major QTL on chromosome 4 for root Na^+ and Na^+:K^+

To identify loci associated with salt tolerance-related phenotypes, GWA mapping was conducted using 397,812 SNPs and eight salinity-related phenotypes collected on 365 rice accessions (Fig 2; Supplemental Fig S1-S4) (McCouch et al., 2016). A linear mixed model

implemented in EMMA was used for the association analysis (Kang et al., 2008). A total of 90 highly significant QTL (245 SNPs; $p < 10^{-5}$) were identified for salinity tolerance traits with the strongest associations detected for root Na^+ content followed by root Na^+:K^+ (Fig 2A and B, respectively). A region located at ~30.6 Mb on chromosome 4 was found to have the largest effect and explained 15% of the phenotypic variation beyond that explained by population structure for root Na^+:K^+ (Supplemental File S2). An additional 25% of the phenotypic variance for root Na^+:K^+ was explained by population structure suggesting that this trait may be heavily influenced by the differences between the major subpopulations in rice.

For each trait, the number of significant QTL ranged from 3-24, with the highest number of QTL identified for root biomass ratio (24 QTL). Many of these QTL had small effects, explaining ~4.7-7.5% of phenotypic variation for root growth response. These results indicate a polygenetic architecture for root growth responses to salinity. A large number of QTL with minor effects (explaining < 7% phenotypic variation) were identified for shoot Na^+ content and Na^+:K^+, suggesting a polygenic architecture for these traits in rice. This trend was observed for all traits, with the exception of root Na^+ and Na^+:K^+, suggesting that salinity tolerance in terms of growth and shoot ion homeostasis in rice is regulated by many loci with small effects. Twenty QTL were commonly detected for two or more traits. Shoot Na^+ and Na^+:K^+ showed the largest number of shared QTL (12 QTL), however much of this similarity is likely driven by the strong phenotypic and genetic correlation observed between these traits within tissues (Table I; Supplemental Table S2).

2.2.3 HKT1;1 Regulates Na^+ homeostasis during the early tillering stage

The most significant QTL for root Na^+ content and Na^+:K^+, named *Root Na^+ Content 4* (*RNC4*), spans a region of ~575 Kb (30,481,871 – 31,057,205) on chromosome 4 (Fig 3A). To characterize this region further and identify candidate genes that may be underlying natural variation for this trait, this region was segmented into haplotype blocks and the contributions of each block to root Na^+ content and Na^+:K^+ were determined using ANOVA. A total of 36 blocks were identified in this 575 Kb region (Supplemental Figs S5-S8). A single 9.7 Kb block from 30,727,920 - 30,737,580 bp was found to have the largest contribution to root Na^+ and Na^+:K^+, with approximately 16% of phenotypic variation explained for root Na^+ and 17.5% explained for root Na^+:K^+ (Fig 3B; Table II). The region spanning from the 5' boundary of block 2 to the 5' boundary of block 3 harbored only two genes, both of which were annotated as sodium transporters, *HKT1;1* and *HKT1;4* (LOC_Os04g51820 and LOC_Os04g51830 respectively; Fig 3C).

To further characterize *HKT1;1* and *HKT1;4*, the expression patterns of both genes were examined in ten tissues at two developmental time points (early seedling and anthesis). The expression of both *HKT1;1* and *HKT1;4* were higher in leaf tissue compared to root tissue during the seedling stage (Fig 4, Fig S9). However, the expression of *HKT1;1* and *HKT1;4* within aerial tissues differed across developmental stages. *HKT1;1* was highly expressed in the leaf blade and leaf sheath during the early seedling stage (Fig 4A). *HKT1;4*, on the other hand displayed the highest expression during reproductive stage, specifically in culm tissue at ~7 days after anthesis (Fig 4B).

To examine whether transcript abundance may be a component of the phenotypic differences observed between allelic groups at *RNC4*, RNA sequencing was performed on shoot tissue of 32 accessions in control and saline conditions, and the expression of both genes was compared between allelic groups at *RNC4* (Fig 4C, D; Supplemental File S3). For both genes, accessions that showed higher root Na^+ content (T allele at SNP-4-30535352), also showed higher expression in both control and saline conditions compared to accessions with low root Na^+ content (G allele at SNP-4-30535352). The expression of *HKT1;1* was approximately 92% higher in high root Na^+ lines in control conditions compared to low root Na^+ lines, while a 44% higher expression was observed in saline conditions (Fig 4C). While the overall expression level was much lower for *HKT1;4* compared to *HKT1;1*, a similar trend in gene expression was also observed between the two allelic groups of *HKT1;4* (Fig 4D). A 46% and 57% higher expression was observed in lines with high root Na^+ content compared to lines with low root Na^+ content in control and saline conditions, respectively (Fig 4D). These results suggest that differences in expression of *HKT1;1* and/or *HKT1;4* may be a component underlying variation in root Na^+ content at *RNC4*.

To determine if these two HKTs within *RNC4* regulate Na^+ content during salinity stress at the early tillering stage, three independent RNA-interference (*RNAi*) lines were generated for both genes. Transcript levels in the leaf tissue was reduced by approximately 2.9-6.2 and 2-2.2 fold in $HKT1;4_{RNAi}$ and $HKT1;1_{RNAi}$ lines compared to wild-type (*WT*) 'Kitaake', respectively (Supplemental Fig S10). A 9 $dS \cdot m^{-1}$ (~90 mM NaCl) was gradually imposed at 10 DAT for 14 days to replicate the stress treatment for the large-scale screening. Reduced expression of *HKT1;1* had severe phenotypic effects on shoot and root ion homeostasis as well as shoot and root growth under salinity. Shoot Na^+ and Na^+:K^+ were 31-41% and 27-41% higher,

respectively, in *HKT1;1*$_{RNAi}$ lines compared to *WT* ($p < 0.0001$, $p < 0.05$ respectively; Fig 5A-C) . A 21-27% reduction in root Na^+ was observed in *HKT1;1*$_{RNAi}$ and 31-33% lower root Na^+:K^+ was observed in *HKT1;1*$_{RNAi}$ compared to *WT* ($p < 0.05$ and $p < 0.0001$, respectively; Fig 5D-F). In *RNAi* plants, shoot and root growth was reduced by 44-55% and 78-72% respectively in salt treated plants relative to those in control conditions, while in *WT* a 26% and 45% reduction in shoot and root growth, respectively was observed in *WT* plants (Supplemental Fig S11). No differences were observed between *HKT1;4*$_{RNAi}$ and *WT* plants (Fig 5, Supplemental Fig S11). These results suggest that *HKT1;1* may influence the shoot and root Na^+ content during the early tillering stage, and is likely the causal gene underlying *RNC4*.

2.2.4 Allelic variants of HKT1;1 alter Na^+ transport

To determine whether there were sequence differences between allelic groups at *RNC4*, sequencing data was mined for variants in *HKT1;1* (Supplemental File S4). Nine variants were detected in the coding region of *HKT1;1* with four SNPs resulting in non-synonymous amino acid substitutions in *HKT1;1* (Fig 6; Fig S12) (Jabnoune et al., 2009). Of the nine variants, only M4 displayed a significant deviation from the expected frequency in the minor allelic group, indicating that it is unlikely to be important for the high root Na^+ phenotype exhibited by accessions in the minor allelic group (Pearson's chi squared test, $p < 1.26 \times 10^{-5}$). The remaining three non-synonymous mutations (M3, M5 and M8) were detected in thirteen accessions all belonging to the minor allelic group, which is characterized by high root Na^+ content, at the most significant SNP for root Na^+ content (SNP-4-30535352). The higher frequency of these three non-synonymous mutations observed in minor allele accessions (T) suggests that allelic variation in *HKT1;1* could be a component in the genetic basis of the observed difference in root Na^+

content between major and minor alleles. No sequence differences in *HKT1;4* were observed between allelic groups at *RNC4*.

To characterize the biophysical properties of the two major isoforms identified between allelic groups at *RNC4*, *HKT1;1* was isolated from two representative accessions, 'Nipponbare' and 'Zhenshan 2', which have the reference and the three non-synonymous mutations at the three locations (M3, M5 and M8), respectively (Fig S12). At the transporter structure level, two non-synonymous SNPs (M8 and M5) lead to amino acid substitutions in cytosolic regions of HKT1;1: proline to leucine within the N-terminal cytosolic region, phenylalanine to serine in the cytosolic loop between the first and second transmembrane segment-pore region-transmembrane segment (MPM) domains (Fig 6A, B; Fig S12). The third non-synonymous SNP results in an asparagine to serine substitution in the external part of the pore-forming region of the second MPM (Fig 6A, B; Fig S12). Functional analysis was performed by voltage-clamp electrophysiology using *Xenopus* oocytes for the two variants of HKT1;1 (Fig 6). The amount of expressed transporters targeted to the oocyte membrane was similar for the two variants, as indicated by the mean GFP fluorescence intensity emitted by either of the tagged transporters at the membrane (Fig 6 C, D). In agreement with previous reports, both isoforms of HKT1;1 displayed low affinity, high Na^+ *versus* K^+ selectivity, inward rectifying activity and no time-dependent kinetics (Fig 6E, Figs S13 and S14) (Jabnoune et al., 2009). However, the two allelic variants displayed considerable differences in Na^+ transport activity. The variant from the accessions with high root Na^+, HKT1;1-Zh, exhibited higher inward (negative) currents compared to that from 'Nipponbare', HKT1;1-Ni (Fig 6E, F), essentially due to a less negative voltage threshold of inward rectifying current activation by 20-25 mV in all ionic conditions (Fig 6E, F, Figs S12 and S13). This latter feature was especially expected to favor transport activity

of HKT1;1-Zh compared to HKT1;1-Ni during salinity stress where the high concentration of Na^+ in the apoplast results in a depolarization of the plasma membrane (Laurie et al., 2002; Mian et al., 2011). Thus, at a weak negative voltage, the current could be more than six-fold higher in HKT1;1-Zh, compared to HKT1;1-Ni (Fig 6G).

To determine if these differences in transport activity have physiological effects *in vivo*, native overexpression lines were generated for each variant (*HKT1;1*$_{Ni}$, *HKT1;1*$_{Zh}$). A ~4.3 kb genomic region was isolated from 'Nipponbare' and 'Zhenshan 2', which included the entire CDS of *HKT1;1* and a 1.9 kb promoter, and was expressed in 'Kitaake'. The endogenous *HKT1;1* in 'Kitaake', at the protein level, is identical to HKT1;1-Ni, and thus lacks the three non-synonymous variants. Two independent transformants for each variant (*HKT1;1*$_{Ni}$, *HKT1;1*$_{Zh}$), each containing only a single copy of the transgene, were evaluated under a 9 $dS \cdot m^{-1}$ salt stress for a period of two-weeks. The expression of *HKT1;1*$_{Zh}$ resulted in an increase in root Na^+ and Na^+:K^+ compared to *HKT1;1*$_{Ni}$, while no differences were observed between variants for root K^+ (Fig 7). A considerable increase in both root Na^+ and Na^+:K^+, as well as a reduction in root K^+ was observed in both native overexpression lines (*HKT1;1*$_{Ni}$ and *HKT1;1*$_{Zh}$) compared to 'Kitaake', which is opposite to the root phenotype observed in the *HKT1;1*$_{RNAi}$ lines (Fig 7). However, expression under the native promoter had no effects on shoot Na^+ or Na^+:K^+ (Fig 7). Together, these results provide further evidence that *HKT1;1* is responsible for the higher root Na^+ phenotype, and that the difference in Na^+ content between the allelic groups at *RNC4* is likely due to functional differences in Na^+ transport by *HKT1;1* alleles, with the three non-synonymous SNPs in *HKT1;1-Zh* resulting in higher Na^+ transport activity.

2.2.5 Origins of HKT1;1 variants in cultivated rice

A difference in allele frequencies of the three non-synonymous mutations in HKT1;1 was observed between the major subpopulations in the 32 sequenced accessions of RDP1. However, since it was difficult to examine subpopulation differentiation with this small of a sample size, the differences were explored in more depth using resequencing data from a larger diversity panel of 3,024 accessions (Alexandrov et al., 2015). A total of 206 SNPs spanning a ~38 Kb region around *HKT1;1* was used for haplotype analysis. In agreement with the allele frequency observed in the 32 accessions of RDP1, a clear differentiation could be observed between *indica* (*ind1A*, *ind1B*, *ind2*, *ind3* and *indx*) and *japonica* (*temp*, *trop1*, *trop2* and *japx*) subspecies in the larger diversity panel (Fig 8). Haplotypes H1, H5 and H8 harbored the three non-synonymous alleles and were found in nearly 85% of the *indica* accessions. The sequence similarity between high root Na^+ haplotypes was very high, ranging from ~88-94% identity. Haplotypes containing high root Na^+ alleles of *HKT1;1* were also found in the *japonica* (*temp*, *trop1*, *trop2* and *japx*), *aus* and *aromatic* subpopulations, albeit at a much lower frequency (0-3%). In contrast, haplotypes H2, H3, H4, and H7 were found predominately in the *japonica* accessions and lacked the high root Na^+ allelic form of *HKT1;1*. Within the low root Na^+ group, haplotypes exhibited high sequence similarity (~65-94%). Given the clear divergence between *indica* and *japonica* for *HKT1;1* haplotypes and the effects of HKT1;1 isoforms on root Na^+ content, collectively these results strongly suggest that a significant proportion of the difference between rice subpopulations in root Na^+ in RDP1 is due to differences in frequency of HKT1;1 variants.

Given the contrasting haplotype frequencies of high and low root Na^+ variants of HKT1;1 between subpopulations of cultivated rice, we explored the origins of these haplotypes by examining their frequencies in a collection of 446 *Oryza rufipogon* accessions collected

throughout South and Southeast Asia (Huang et al., 2012). These accessions represent three major populations (*Or-I*, *Or-II* and *Or-III*) and provide an adequate representation of the ancestral populations of cultivated rice (Huang et al., 2012). Two haplotypes (H1 and H5) were identified that harbored the high root Na^+ variants of *HKT1;1*, and were found in nearly 70% of the *O. rufipogon* accessions. The H1 haplotype displayed the highest frequency in the *Or-II* clade and was also found in the majority of *indica* accessions, suggesting that the *indica* allele is likely derived from *Or-II*. In contrast, two haplotypes (H2 and H6) were identified with the low root Na^+ variant and were present in only 19% of the *O. rufipogon* accessions. The H6 haplotype was the most frequent and present in 18% of the *O. rufipogon* accessions, but absent from the *japonica* cultivated rice accessions. In contrast, H2 occurred at high frequency (44%) in cultivated *japonica*, particularly the *tropical japonica* subpopulation, suggesting that H2 is potentially the ancestral haplotype for the *japonica* subspecies. Interestingly, the haplotypes found at high frequencies in the *japonica* subspecies were present at considerably lower frequencies in wild rice accessions (the highest frequency observed was 0.16), indicating that these haplotypes in *japonica* subspecies may be derived from a relatively small population of wild progenitors.

2.3 Discussion

Salinity tolerance is a complex polygenic trait and is regulated by several physiological mechanisms (Munns and Tester, 2008). Salinity reduces plant growth through osmotic effects, which are experienced shortly after the addition of Na^+ to the external media, and ionic effects, which are experienced later in the stress as Na^+ accumulates in the leaves to toxic levels. The ability to maintain growth in saline conditions involves a suite of physiological mechanisms including osmotic adjustment, the exclusion of sodium from leaf tissues by sequestration in the

root or leaf sheath, the storage of Na^+ into vacuoles or partitioning in tissues where the toxic effect of Na^+ is reduced (Munns and Tester, 2008). In this study, the complex polygenic nature of salinity tolerance in rice is evidenced by the large number of loci with small effects identified for shoot and root growth responses.

Although shoot Na^+ exclusion is often used as a parameter for salt tolerance, the relationship between low shoot Na^+ and the ability to maintain growth in saline conditions does not always hold true (El-Hendawy et al., 2005; Genc et al., 2007). Here, no significant relationships were observed between ion traits and growth responses across all the subpopulations, suggesting that in the current experimental conditions other tolerance mechanisms besides Na^+ exclusion may be important for salt tolerance in rice. In low to moderate salinity, the osmotic effects of high Na^+ in the external media are likely to have a much greater impact on plant growth, compared to ionic effects (Munns and Tester, 2008). During the ionic phase of salt stress, Na^+ must accumulate to toxic levels to cause cell death and impede growth. Thus, ionic tolerance mechanisms begin to play a role much later in the stress, or when the concentration of Na^+ in the external media is high. Other studies that have exposed diverse rice accessions to higher concentrations of NaCl and/or for longer periods have reported weak to moderate relationships between ion traits and growth responses to salinity (Lee et al., 2003; Negrão et al., 2013; Platten et al., 2013). Thus, Na^+ exclusion may be important during more severe stress treatments than was used in the current study. The relatively moderate salt stress imposed in the current study may not be enough for Na^+ to accumulate to toxic levels to significantly inhibit growth, and may partially explain the lack of correlation, both phenotypic and genetic, between ion traits and growth responses.

2.3.1 HKT1;1 regulates root Na^+ content

RNC4 harbors two Na^+ transporter genes, *HKT1;1* and *HKT1;4*. *HKTs* are well-known components of salinity tolerance in several plant species including rice (*HKT1;5* is likely the causal gene in the *SalTol* QTL), wheat and *Arabidopsis* (Rus et al., 2001; Berthomieu et al., 2003; Ren et al., 2005; Sunarpi et al., 2005; Huang et al., 2006; Byrt et al., 2007; Horie et al., 2009; Munns et al., 2012; Wang et al., 2015; Suzuki et al., 2016). Although both *HKT1;1* and *HKT1;4* displayed significant differences in expression between allelic groups at *RNC4*, several key findings suggest that HKT1;1 is more important for root Na^+ content during the early tillering stage and for the salinity level imposed in our experimental set-up. First, the genes are expressed at different developmental stages. *HKT1;1* was expressed at the highest levels in blade and leaf sheath tissues of seedlings, while *HKT1;4* showed the highest expression in culms of mature plants (Fig 4B). Second, reduced expression of *HKT1;1* in transgenic *RNAi* lines resulted in a greater sensitivity to salinity compared to *WT*, while $HKT1;4_{RNAi}$ and *WT* plants displayed similar phenotypes under salinity (Fig 5). In a recent report, Suzuki *et al* showed that *HKT1;4* is primarily expressed in peduncles during flowering (14 week old plants) and, through *RNAi*, showed that *HKT1;4* is primarily involved in Na^+ homeostasis only during the reproductive phase (Suzuki et al., 2016). Since the current study was conducted during the early tillering stage (< 1 month old plants), it is unlikely that this gene would have an impact on salinity tolerance in this developmental window. Finally, increased expression of *HKT1;1* with the native promoter resulted in higher Na^+ in root tissue, which is identical to the phenotype associated with *RNC4*. Together, these data suggests that *HKT1;1* is the causal gene underlying *RNC4* and contributes to root Na^+ content during the early tillering stage.

2.3.2 Variants in HKT1;1 underlie natural variation for root Na^+ content

The differences in Na^+ content observed between allelic groups at *RNC4* is likely due to functional differences in Na^+ transport by *HKT1;1* alleles, with the three non-synonymous SNPs in *HKT1;1-Zh* resulting in higher Na^+ transport activity. Na^+ transport occurred at less negative voltages in the isoform found in accessions with high root Na^+ compared to that isolated from accessions with low root Na^+. During salt stress, the accumulation of Na^+ in the apoplastic space increases HKT1;1 Na^+ transport activity, the apparent affinity for Na^+ of this transporter type is particularly low (Km ~ 80 mM; Fig S12), but in the meantime, uptake of Na^+ from the apoplast results in membrane depolarization, which reduces HKT1;1 conductance due to inward rectification property (Jabnoune et al., 2009). In the high root Na^+ isoform, a higher (less negative) voltage threshold of current activation was observed, for instance in the presence of 10 mM external Na^+ noticeable Na^+ transport was observed between -75 and -90 mV, while in the low root Na^+ isoforms, activation occurred at more negative voltages (Fig 6 C, D). Thus, lower Na^+ concentrations are required to induce Na^+ uptake in the high root Na^+ isoform of HKT1;1. In summary, the enhanced ability to transport Na^+ in accessions harboring the high root Na^+ isoform of HKT1;1 is likely due to the early activation of Na^+ transport.

2.3.3 HKT1;1 isoforms are derived from independent populations during domestication

Indica varieties have long been recognized to as a source of salt tolerance, largely due to Na^+ exclusion from leaf tissue. The most widely used QTL, *SalTol*, was identified by Lin *et al.* using a biparental population derived from the salt tolerant *indica* landrace 'Nona Bokra' and sensitive *japonica* variety 'Koshihikari' (Lin et al., 2004). Tolerance mediated by *SalTol* is associated with the exclusion of Na^+ from shoot tissue, through the removal of Na^+ from the xylem and sequestration in xylem parenchyma cells in the root tissue (Ren et al., 2005; Thomson

et al., 2010). While several studies have demonstrated that the *indica* subspecies harbors many varieties exhibiting high shoot Na^+ exclusion ability, tolerant alleles in *SalTol* have only been utilized from a few *indica* landraces, and it is likely that other loci are contributing to Na^+ exclusion in the *indica* subspecies (Yeo et al., 1990; Gregorio et al., 2002).

In agreement with previous studies, a considerable difference among the five subpopulations was observed in root and shoot Na^+ content and Na^+:K^+, with *indica* accessions generally displaying higher root Na^+ content and Na^+:K^+, as well as slightly lower shoot Na^+ and Na^+:K^+. The relationship between root and shoot ion traits (specifically Na^+ and Na^+:K^+) differed considerably within each of the subpopulations. For instance, positive correlations were observed between tissues for Na^+ and Na^+:K^+ in the *tej*, *trj* and *aus* subpopulations. However, in the *indica* and *admix* subpopulations no relationships were observed between tissues for Na^+ and Na^+:K^+. The moderate positive genetic correlation observed between tissues across all accessions of RDP1 indicates that these traits may be regulated in part by common genes. However, this may be highly dependent on the subpopulation. The high frequency of the Na^+ accumulating isoform for of HKT1;1 in the *indica* and *admix* subpopulations may "uncouple" the relationship between tissues for Na^+ and Na^+:K^+.

The contrasting root Na^+ content observed between *indica* and *japonica* accessions of RDP1 is consistent with the differences in transport activity and the frequencies of the high and low root Na^+ isoforms of *HKT1;1*. The haplotypes of *HKT1;1* could be clearly separated into two distinct groups, corresponding to the *japonica* (H2, H3, H4 and H7) and *indica* predominate forms (H1 and H5). The high root Na^+ haplotypes (H1, H5 and H8) were most frequent in *Oryza rufipogon*, while the low root Na^+ haplotypes were identified in only ~31% of the *Oryza*

rufipogon accessions and were nearly fixed in *japonica* accessions. The two major subspecies of *Oryza sativa* were domesticated from two geographically isolated populations of *Oryza rufipogon* (Sweeney and McCouch, 2007; Huang et al., 2012). The low diversity in *japonica* germplasm reported by several studies is consistent with a bottleneck during domestication, and suggests that the *japonica* subspecies may be derived from a relatively small founding population of *Oryza rufipogon* (Caicedo et al., 2007; Mather et al., 2007; Huang et al., 2010, 2012) (Supplemental Fig S15). Although the high root Na^+ isoform was found in ~30% of the Or-III subpopulation, the founding subpopulation of *Oryza rufipogon*, it is plausible that the bottleneck experienced during domestication may have resulted in the loss of the high root Na^+ *HKT1;1* variant from *japonica* subspecies.

2.3.4 The role of HKT1;1 in Na^+ exclusion from shoot tissue

Like many other *HKT* members, *HKT1;1* is well-expressed in the vascular tissue of the shoot, and to a lesser extent in the root (Horie et al., 2009; Jabnoune et al., 2009; Wang et al., 2015). In the current study, $HKT1;1_{RNAi}$ lines were more sensitive to salt stress, and exhibited higher shoot Na^+ content and lower root Na^+ content compared to *WT* plants. The expression patterns of *HKT1;1*, as well as the phenotypes exhibited by $HKT1;1_{RNAi}$ lines are in agreement with those reported by Mäser *et al* for *AtHKT1;1* in *Arabidopsis*, suggesting that the genes may have similar physiological functions (Mäser et al., 2002). Like $HKT1;1_{RNAi}$, *athkt1;1* knockout mutants are hypersensitive to salt stress and exhibit higher shoot Na^+ and lower root Na^+ (Mäser et al., 2002; Davenport et al., 2007). In rice, Wang *et al* showed that *hkt1;1* knockout mutants accumulate Na^+ in xylem sap and display a reduction in Na^+ in phloem sap compared to *WT* (Wang et al., 2015). These observations together with the observed accumulation of Na^+ in shoot tissue prompted Wang *et al* to suggest that *HKT1;1* may regulate sodium exclusion from the

shoot of seedlings possibly through xylem-to-phloem or parenchyma-to-xylem transfer of Na^+ (Wang et al., 2015). Such xylem-to-phloem transfer of Na^+ by a HKT member has been debated in *Arabidopsis* (Berthomieu et al., 2003; Sunarpi et al., 2005; Davenport et al., 2007). In agreement with *hkt1;1* mutant phenotype reported by Wang *et al*, *athkt1;1* knockout mutants also exhibit higher xylem Na^+ and lower phloem Na^+ (Berthomieu et al., 2003; Sunarpi et al., 2005; Wang et al., 2015). Although *AtHKT1;1* was initially proposed to function in the recirculation of Na^+ from the root to the shoot (via loading of Na^+ into the phloem in the shoots), Sunarpi *et al* later proposed that *AtHKT1;1* functions primarily in the removal of Na^+ from the xylem sap and eventually to the phloem through symplastic diffusion (Berthomieu et al., 2003; Sunarpi et al., 2005). However, a later study showed that *AtHKT1;1* was primarily involved in the retrieval of Na^+ from the xylem in root tissue, and suggested that the function of AtHKT1;1 in shoot tissue may be dependent on the experimental conditions (discussed in (Munns and Tester, 2008)) (Davenport et al., 2007). For the case of *HKT1;1* in rice, further studies are required to provide the exact mechanism for the regulation of root Na^+ content and/or shoot Na^+ exclusion, and are outside the scope of this manuscript.

Given the phenotypes exhibited by $HKT1;1_{RNAi}$ lines, as well as the proposed function described by Wang *et al.*, the absence of an association of *HKT1;1* with shoot Na^+ or Na^+:K^+ is surprising (Wang et al., 2015). If *HKT1;1* regulates retrieval of Na^+ from the parenchyma or xylem in shoot tissues, one would expect that the high root Na^+ allele would also have a large impact on shoot Na^+ content. However, the concentration of Na^+ in shoot tissue is likely more dependent on the amount of Na^+ loaded into the xylem, and thus mechanisms which limit the delivery of Na^+ to xylem stream would likely be more effective mechanism for shoot Na^+ exclusion (Munns and Tester, 2008). Without an effective mechanism to limit Na^+ entry into the

xylem stream in the root, very high expression of *HKT1;1*, or a highly active variant of HKT1;1 would likely be necessary to reduce shoot Na^+ content. While the *indica* (high root Na^+ content) variant of HKT1;1 displayed higher transport activity compared to *japonica* variant (low root Na content), it is likely that these biophysical differences are not sufficient to have an impact on shoot Na^+ content.

Other members of the *HKT* family have been identified that are expressed in the vascular tissue of the root, and primarily function to remove Na^+ from the xylem to limit the delivery of Na^+ to the shoot. In rice, this function is largely achieved through the action of *HKT1;5* (Ren et al., 2005; Cotsaftis et al., 2012). In contrast to *HKT1;1*, *HKT1;5* is mostly expressed in the root and therefore is essentially involved in xylem sap desalinization (Ren et al., 2005). In the current study, the *SalTol* QTL that harbors *SKC1/HKT1;5* explained only a small portion of phenotypic variation for shoot Na^+ and shoot Na^+:K^+ (~6%; SNP-1.11472400). Several studies have identified alleles within *SKC1/HKT1;5* that are associated with Na^+ exclusion and salt tolerance, but it is unclear whether the effects of these alleles are as strong as those reported by Gregorio *et al.* and Bonilla *et al.* (Gregorio, 1997; Bonilla et al., 2002; Lin et al., 2004; Ren et al., 2005; Thomson et al., 2010; Bandillo et al., 2013). Given the small effect of this QTL in the current study, as well as the large number of QTL identified for shoot Na^+ and Na^+:K^+, it is likely that natural variation for shoot Na^+ and Na^+:K^+ involves additional genetic components in addition to *SKC1/HKT1;5*.

2.4 Materials and Methods

2.4.1 Plant materials and genotyping

This study included 383 of the 421 original RDP1 accessions, as well as seven check varieties (Zhao et al., 2011; Famoso et al., 2011; Eizenga et al., 2014). Accessions were obtained from the USDA-ARS Dale Bumpers Rice Research Center and purified through single seed descent before they were phenotyped. Thirty-eight accessions of RDP1 were not included because of lack of seed availability and/or poor seed quality. The set of accessions from RDP1 included 77 *indica*, 52 *aus*, 92 *temperate japonica*, 85 *tropical japonica*, 12 *groupV/aromatic*, and 56 highly admixed accessions (nine accessions were unassigned), according to the classification by Famoso *et al* (Famoso et al., 2011). A total of 365 accessions from RDP1 were genotyped using 700,000 SNPs (McCouch et al., 2016). Filtering SNPs based on minor-allele frequency (MAF > 0.05) left ~397,812 high quality SNPs (depending on the trait analyzed) (McCouch et al., 2016). Previous results indicated LD decays to 0.20 between 0.5-1.0 Mb, indicating the marker density provided by the SNP array has suitable power to detect linked causal variants of moderate to large effect QTL (McCouch et al., 2016).

2.4.2 Growth conditions and salt treatment

The experiment was conducted between July to Sep 2013 in a controlled green house at Lincoln, NE. Rice (*Oryza sativa*) seeds were dehusked manually and germinated in the dark for two days at 28°C in a growth cabinet (Percival Scientific). Twelve hours before transplanting seeds were exposed to light (120 μmol m^{-2} s^{-1}). The green house conditions were as follows: photoperiod (16:8 day:night), temperature 25-28 °C and humidity 50-80%. Seedlings were transplanted into the pots filled with Turface (Profile Products, LLC) and were grown in tap

water for four days after transplanting. For the remainder of the experiment the plants were supplemented with half strength Yoshida solution (pH 5.8) (Yoshida et al., 1976). Salt treatment was applied as described previously by Walia *et al.* with minor modifications (Walia et al., 2005). Briefly, NaCl was mixed with $CaCl_2$ in a 6:1 molar ratio and was added after 10 d of seedling growth. The stress treatment was started at 2.5 $dS \cdot m^{-1}$ which increased gradually up to 9.5 $dS \cdot m^{-1}$ in 4 steps over a period of four days (~2 $dS \cdot m^{-1}$ or 20 mM NaCl per day) to avoid any osmotic shock to the plants. The stress treatment was stabilized at 9.5 $dS \cdot m^{-1}$ for next two weeks. The nutrient solution pH and electrical conductivity (EC) were monitored and maintained twice daily. The pH of the nutrient solution was maintained at 5.8 using H_2SO_4 and KOH. Root and shoot samples were collected separately and rinsed 3 times in tap water and once in deionized water to remove excess NaCl at the completion of the experiment (14 days of 9.5 $dS \cdot m^{-1}$; 28 days after transplant). The samples were oven dried at 60 °C for one week prior to measuring root and shoot biomass. Shoot and roots from two plants were taken for biomass measurement.

2.4.3 Ion content measurement

For the large-scale screening of RDP1 dried shoot samples were ground and 200 – 300 g of total material was digested with 0.1N Nitric acid (Fisher Scientific) at 70 °C for 8 hrs. Root samples were weighed and digested without any grinding. Samples were diluted and cation (Na^+ and K^+) concentrations in the plant extract were determined with appropriate standard by dual Flame photometry (Cole Parmer, USA).

2.4.4 Statistical analysis of phenotypic data

Data was combed across periods and a linear model was fit to calculate adjusted means for individual accession using the PROC GLM procedure of the Statistical Analysis System

(SAS Institute, Inc.). The linear model included period (i.e., June-July or Aug-Sept), replication nested within period, tub nested within replication, accession, and accession-by-period interaction.

For the purpose of estimating variance components, a second similar linear model was fit using PROC MIXED in SAS. This time, all effects were assumed to be random effects. Broad-sense heritability (H^2) on an entry-mean basis was calculated as $H^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_{GP}^2 / 2 + \sigma_e^2 / 6)$

Where σ_G^2 is the variance among accessions, σ_{GP}^2 is the accession-by-period interaction variance, and σ_e^2 is the error variance. In this context, the divisor 2 is equal to the number of periods and the divisor 6 is equal the number of replications per period (three) multiplied by the number of periods. Broad-sense heritability provides a sense of how much of the total variation observed is due to genetic variation among accession, and indicates the power of GWAS.

2.4.5 Mixed linear model for genome-wide association analysis

Marker-trait associations were tested using the linear mixed model y= Xβ + Cγ + Zu + e where y is a vector of phenotype; β is a vector of fixed marker effects; γ is a vector of principal component (PC) effects fit in order to account for population structure; u is a vector of polygenic effects caused by relatedness; e is a vector of residuals; X is a marker incidence matrix relating β to y; C is an incidence matrix relating γ to y which consists of the first four principal components (PCs) resulting from a PC analysis; Z is the corresponding design matrix relating y to u. It is assumed $\mathbf{u} \sim MVN(0, \mathbf{K}\sigma_u^2)$ and $\mathbf{e} \sim MVN(0, \mathbf{I}\sigma_e^2)$ where K is a standardized kinship matrix estimated using an allele-sharing matrix calculated from the SNP data. The above model was

implemented using the efficient mixed-model association (EMMA) algorithm of Kang *et al* (Kang et al., 2008).

The method published by Li and Ji was used to determine a comparison-wise error rate to control the experiment-wise error rate (Li and Ji, 2005). Briefly, the correlation matrix and eigenvalue decomposition among 397,812 SNPs were calculated to determine effective number of independent tests (M_{eff}). The test criteria was then adjusted using the M_{eff} with the Sidak correction below

$$\alpha_p = 1 - (1- \alpha_e)^{1/Meff},$$

where α_p is the comparison-wise error rate and α_e is the experiment-wise error rate (Šidák, 1967). An $\alpha_e = 0.05$ was used in this study.

Analysis of variance (ANOVA) was used to estimate proportion of phenotypic variance accounted for by significant SNPs after adjusting for population structure effects. A 200 kb window was used to define groups of significant SNPs tagging the same locus. Only the most significant SNP within a 200 kb window was used to tag that locus. The percent variation explained by each significant SNPs was determined by comparing the linear models, $y= X\beta + C\gamma + e$, and $y= C\gamma + e$, where β is the SNP effect; γ is a vector of PCs effects to account for population structure; X is a vector of SNP genotypes; C is an incidence matrix relating γ to y which consists of the first four principal components (PCs). Therefore, the effect of each SNP is reported after accounting for the effects of population structure.

2.4.6 Estimation of genetic correlation

Genetic correlations between traits were estimated with and without correcting for population structure and family relatedness. The rationale behind correcting genetic correlations for population structure is to measure the correlation independent of long-range LD between loci caused by population structure (Wisser et al., 2011). To accomplish this, a multivariate mixed model was fit as described by Wisser *et al.* including all traits as response variables; fixed experimental design effects (replication and tub nested within replication); fixed population structure effects modeled using the four PCs as above; random polygenic effects modeled using the kinship matrix as in the GWAS model described above; and random residuals assumed to independent and identically distributed (Wisser et al., 2011). Restricted maximum likelihood implemented in ASReml-R v.3.0 was used to estimate genetic and residual variances, and genetic and residual covariances among traits (Gilmour et al., 2009). Estimates of genetic variances and covariances were used to calculate genetic correlations among traits. For estimation of genetic correlations uncorrected for population structure, the same methods were used except population structure and polygenic effects were not included in the mixed linear model.

2.4.7 Haplotype analysis

Haplotype blocks were constructed using the four gamete method (4gamete) implemented in the software Haploview (Barrett et al., 2005). The method creates block boundaries where there is evidence of recombination between adjacent SNPs based on the presence of all four gametic types. We used a cut-off of 2%, meaning that if addition of a SNP to a block resulted in recombinant alleles at a frequency exceeding 2%, the SNP was not included in the block.

2.4.8 Haplotype analysis of HKT1;1 in cultivated rice and Oryza rufipogon

To examine the frequency of high and low root Na^+ forms of HKT1;1 in a set of 3,023 cultivated rice and 446 *Oryza rufipogon* accessions, a set of 206 SNPs was extracted from a ~37 kb (30,700,524 – 30,737,580) region on chromosome 4. Sequence data for the cultivated rice was obtained from ~9 million genome-wide SNPs generated by the 3000 Rice Genomes Project (3K RGP) (Alexandrov et al., 2015). The 206 SNPs for *Oryza rufipogon* was obtained from riceHap3 (www.ncgr.ac.cn/ricehap3/) (Huang et al., 2012). Since SNPs were mapped to different genome builds (IRGSP4.0 to IRGSP1.0 for 3kg and RiceHap3, respectively), the coordinates were converted by aligning a 37 kb region from IRGSP4.0 to IRGSP1.0 using BLAT (Kent, 2002).

Haplotype block analysis was performed using the 4Gamete rule, with a cutoff of 1% in Haploview (Barrett et al., 2005). The frequency of each haplotype within in each subpopulation was determined in R. A haplotype network for this 37 kb region was built with PopArt (Leigh and Bryant, 2015). Nucleotide diversity (π) was determined at each position for *indica* (*ind1A*, *ind1B*, *ind2*, *ind3* and *indx*), *japonica* (*temp*, *trop 1*, *trop 2*, and *japx*) and wild rice using the "site-pi" function in VCFtools (Danecek et al., 2011).

2.4.9 Growth conditions and sample collection for RNA sequencing

For gene expression analysis, plants were grown in a controlled environment growth chamber. Temperatures were maintained at 28 °C and 25 °C in day and night respectively, relative humidity was maintained at 60% in both day and night. Lighting was maintained at 800 $\mu moles \cdot m^{-2} \cdot s^{-1}$ using high pressure sodium lights (Phillips). Seeds preparation and salt treatment were performed as described above. Eight day (four days after transplant) old rice seedlings were

subjected to 6 $dS \cdot m^{-1}$ for a period of 24h. Salinity stress was increased to 6 $dS \cdot m^{-1}$ gradually in two 3 $dS \cdot m^{-1}$ intervals over a period of 24h. After 24h of 6 $dS \cdot m^{-1}$, aerial parts of the seedlings were excised from the roots and frozen immediately in liquid nitrogen. The samples were ground with Tissuelyser II (Invitrogen) and total RNA was isolated with RNAeasy isolation kit (Qiagen) according to manufacturer's instructions. On-column DNAse treatment was performed to remove genomic DNA contamination (Qiagen). Sequencing was performed using Illumina HiSeq 2500. Sixteen cDNA libraries were combined in each lane.

2.4.10 RNA-seq mapping and quantification

After being examined using the package FastQC, short reads, obtained from Illumina 101-bp single-end RNA sequencing, were screened and trimmed using Trimmomatic to ensure each read has average quality score larger than 30 and longer than 15 bp (Andrews, 2010; Bolger et al., 2014). The trimmed short reads were mapped against to the rice genome (*Oryza sativa* MSU Release 6.0) using TopHat (v.2.0.10), allowing up to two base mismatches per read. Reads mapped to multiple locations were discarded (Trapnell et al., 2009). Numbers of reads in genes were counted by the HTSeq-count tool using gene annotations for the same version of rice genome and the "union" resolution mode was used (Anders et al., 2014).

2.4.11 Variant identification

For a given genotype, all mapped RNA-seq short reads were sorted and indexed by Samtools (Version: 0.1.18) (Li et al., 2009). Single nucleotide polymorphisms (SNPs) and small insertions/deletions (Indels) were identified based on differences between short reads from the given genotype and the reference genome sequence with mapping quality larger than 25, read

depth more than 30, but less than 500. Variations in regions of interest in the rice genome were selected with their coordinates and gene annotations.

2.4.12 First strand cDNA synthesis and real-time quantitative PCR

First strand cDNA synthesis for real-time quantitative PCR (qRT-PCR) was performed using iScript Reverse Transcription Supermix (Bio-Rad Laboratories, Inc., Hercules, CA, USA) using 2 µg of total RNA. For the qPCR reaction, 3 µL of the diluted cDNA (1:20) was used in the 15 µL reaction mixture. In the qPCR reaction volume, 7.5 µL of LightCycler® 480 SYBR Green I Mastermix was used (Roche Diagnostics, Indianapolis, IN, USA). The qRT-PCR was carried out using Roche LightCycle 480 II with the following parameter settings (Roche Diagnostics, Indianapolis, IN, USA): 95 ^{o}C pre-incubation for 5 min, amplification was done for 40 cycles at 95 ^{o}C for 20 sec and 60 ^{o}C for 15 sec and extension at 72 ^{o}C for 15 sec; the melting curve was set-up for 95 ^{o}C, 65 ^{o}C, 97 ^{o}C; cooling was set-up at 40 ^{o}C for 30 sec. We used two independent tissue samples, with tissue from two to three plants pooled for each sample. LOC_Os04g02820 was used as an internal reference gene, which displayed stable expression in all samples analyzed. Relative expression was determined using the delta-delta Ct method (Livak and Schmittgen, 2001). Primer sequences are provided as Supplemental Table S9.

2.4.13 Transgene construction and Agrobacterium-mediated transformation

For *HKT1;1*, a 112 bp region was amplified from genomic DNA of the *japonica* rice variety ‘Kitaake’, while a 95 bp region was amplified for *HKT1;4*. The fragment from *HKT1;4* was ligated into the pENTR-D-TOPO vector, while for *HKT1;1* the fragment was inserted into pDONR221 using the BP reaction following the manufacturer’s instructions (Invitrogen). Finally, each fragment was introduced into the pANDA RNAi expression vector (Miki and

Shimamoto, 2004; Miki et al., 2005). Transformation of 'Kitaake' calli was performed according to the methods outlined by Cheng *et al.* using the EHA-105 strain of *Agrobactrium* (Cheng et al., 1998). Calli and plants were selected on ½ strength MS media supplemented with 50 µg/ml hygromycin. The expression of *HKT1;1* and *HKT1;4* in shoot and flag leaf tissue of T1 plants, respectively, was determined using realtime PCR using the same conditions as described above. Primer sequences are provided in Supplemental Table S9.

To generate native overexpression lines for the two isoforms of HKT1;1, a ~4.3 kb fragment was amplified from 'Nipponbare' and 'Zhenshan 2'. The fragments were cloned into pDONR221 vector via the BP reaction, and were subsequently cloned into a pMDC99 backbone with a NOS terminator (Curtis and Grossniklaus, 2003). *Agrobacterium*-mediated transformation and selection of transformants was performed as described above. T1 plants with a single insertion were selected on ½ strength MS media supplemented with 50 µg/ml hygromycin and used for phenotyping.

2.4.14 Evaluation of transgenic plants

Phenotyping of transgenic plants was performed in a controlled environment growth chamber. Three independent RNAi lines (T2) for *HKT1;1* and *HKT1;4* was screened for salinity tolerance, while two independent native overexpression lines (T1 generation) were evaluated for each isoform of *HKT1;1* (*HKT1;1*$_{Ni}$ and *HKT1;1*$_{Zh}$). Temperatures were maintained at 28 °C and 25 °C in day and night respectively, relative humidity was maintained at 60% in day and night. Lighting was maintained at 800 µM using high pressure sodium lights (Phillips). Seeds were surface sterilized in a 40% bleach solution for 20 min, rinsed in sterile water and were germinated on ½ MS media supplemented with 50 ug/ml of hygromycin. *WT* seeds received the

same treatment, but were grown on ½ MS. The seeds were germinated for 24h in complete darkness then were transferred to a growth cabinet (Percival Scientific) and grown for four days at 28°C in 16/8h light (120 μmol m^{-2} s^{-1}). Seedlings were transplanted into the pots filled with Turface (Profile Products, LLC) and were grown in tap water for four days after transplanting. For the remainder of the experiment the plants were supplemented with half strength Yoshida solution (pH 5.8) (Yoshida et al., 1976). Eight days after transplanting a gradual salt stress was applied in three 3 $dS \cdot m^{-1}$ intervals over a period of 24h. The final 9 $dS \cdot m^{-1}$ salt level was maintained for two weeks. Sample collection was performed as described above.

2.4.15 Electrophysiology

To generate constructs for assessing Na^+ transport activities in *Xenopus laevis* oocytes, *HKT1;1* was amplified from cDNA from 'Nipponbare' and 'Zhenshan 2', which are representative accessions for the high and low root Na^+ groups at *RNC4*, respectively, and ligated into the pGEM-Xho vector (Ben Amar et al., 2014). The pGEM-Xho contains the T7 promoter and 5′- and 3′-untranslated regions of the *Xenopus β-globin* gene, which enhances expression in *Xenopus* oocytes. For N-terminal GFP::HKT1;1 fusion constructs, *HKT1;1* was amplified from cDNA from 'Nipponbare' and 'Zhenshan 2' and cloned into pGWB6 using the Gateway LR reaction. GFP::HKT1;1 was then amplified from each construct using primers with SpeI and SalI restriction sites, and introduced into the pGEM-Xho vector (Nakagawa et al., 2007).

Capped and polyadenylated RNA were obtained from linearized vector by *in vitro* transcription, using the mMESSAGE mMACHINE T7 kit (Ambion, USA). Oocytes isolated as previously described were injected with 50 ng of *HKT1;1-Ni* or *HKT1;1-Zh* cRNA (equivalent amount of transporter cRNA in GFP-tagged form) in 50 nL of RNase-free water, or with 50 nL

of RNase-free water (for control oocytes), and then kept for 24 to 48 h at 19 °C in ND96 medium (96 mM NaCl, 2 mM KCl, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$, 2.5 mM sodium pyruvate, and 5 mM HEPES/NaOH, pH 7.4) supplemented with 0.5 $mg \cdot L^{-1}$ of gentamicin, until experiments (Ben Amar et al., 2014). Whole oocyte currents and membrane potential were recorded using the two-electrode voltage-clamp technique with a GeneClamp 500B amplifier (Axon Instruments, USA) 1 to 2 days after cRNA injection. Voltage-pulse protocols, data acquisition and analysis were performed using pClamp9 software (Axon Instruments). Correction was made for voltage drop through the series resistance of the bath and the reference electrode using two external electrodes connected to a bath probe (VG-2A x100 Virtual-ground bath clamp; Axon Instruments). Electrodes were filled with 3 M KCl. The oocytes were continuously perfused during the voltage-clamp experiment with bath solutions containing varying concentrations of monovalent cations (as glutamate or chloride salts) in a background of 6 mM $MgCl_2$, 1.8 mM $CaCl_2$, and 10 mM MES-1,3-bis[tris(hydroxymethyl) methylamino]propane, pH 5.5. The chloride concentration was constant in each set of solutions. D-Mannitol was added when necessary to adjust the osmolarity, which was set to 220–240 mosM in each set of solutions. Voltage-clamp protocol consisted in successive steps of membrane voltage application from -165 to +15 mV in +15 mV increments during 0.5 s, each step beginning with 0.15 s and ending with 0.3 s at the resting potential of the oocyte membrane in the tested bath solution. Mean currents recorded in water-injected control oocytes from the same batch and in the same ionic conditions as HKT-expressing ones were subtracted from those recorded in HKT-expressing oocytes in order to extract HKT-mediated currents from total oocyte currents. HKT1;1-Ni and -Zh current–voltage (I–V) relationships were constructed with transporter extracted currents. The activation potential

of HKT currents was estimated as the lowest voltage at which the current in HKT-expressing oocytes reached twice that in control oocytes.

Confocal observations were made on dark poles of oocytes of similar sizes on a Leica SP8 microscope, using a 20x/0.7dry objective. GFP was excited with a 488 nm laser, and spectral acquisitions of emitted fluorescent light were performed between 495 and 645 nm using a bandwidth of 5 nm, to assert GFP specificity. For each oocyte, mean fluorescence intensity at the membrane was determined from at least 2 optical sections, analyzing 3 ROIs per section using ImageJ (https://imagej.nih.gov/ij/) software.

2.5 References

Ahmadi, N., S. Negrão, D. Katsantonis, J. Frouin, J. Ploux, P. Letourmy, G. Droc, P. Babo, H. Trindade, G. Bruschi, and others. 2011. Targeted association analysis identified japonica rice varieties achieving Na+/K+ homeostasis without the allelic make-up of the salt tolerant indica variety Nona Bokra. Theor. Appl. Genet. 123(6): 881–895.

Alexandrov, N., S. Tai, W. Wang, L. Mansueto, K. Palis, R.R. Fuentes, V.J. Ulat, D. Chebotarov, G. Zhang, Z. Li, R. Mauleon, R.S. Hamilton, and K.L. McNally. 2015. SNP-Seek database of SNPs derived from 3000 rice genomes. Nucleic Acids Res. 43(D1): D1023--D1027.

Ben Amar, S., F. Brini, H. Sentenac, K. Masmoudi, and A.-A. Véry. 2014. Functional characterization in Xenopus oocytes of Na+ transport systems from durum wheat reveals diversity among two HKT1; 4 transporters. J. Exp. Bot. 65(1): 213–222.

Anders, S., P.T. Pyl, and W. Huber. 2014. HTSeq – A Python framework to work with high-throughput sequencing data. Bioinformatics 31(2): 166–169.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Bandillo, N., C. Raghavan, P.A. Muyco, M.A.L. Sevilla, I.T. Lobina, C.J. Dilla-Ermita, C.-W. Tung, S. McCouch, M. Thomson, R. Mauleon, R.K. Singh, G. Gregorio, E. Redoña, and H. Leung. 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice (N. Y). 6: 11.

Barrett, J.C., B. Fry, J. Maller, and M.J. Daly. 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21(2): 263–265.

Berthomieu, P., G. Conéjéro, A. Nublat, W.J. Brackenbury, C. Lambert, C. Savio, N. Uozumi, S. Oiki, K. Yamada, F. Cellier, F. Gosti, T. Simonneau, P. a. Essah, M. Tester, A.A. Véry, H.

Sentenac, and F. Casse. 2003. Functional analysis of AtHKT1 in Arabidopsis shows that Na+ recirculation by the phloem is crucial for salt tolerance. EMBO J. 22(9): 2004–2014.

Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics: btu170.

Bonilla, P., D. Mackell, K. Deal, and G. Gregorio. 2002. RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (Oryza sativa L.) using recombinant inbred lines. Philipp. Agric. Sci. 85: 68–76.

Byrt, C.S., J.D. Platten, W. Spielmeyer, R.A. James, E.S. Lagudah, E.S. Dennis, M. Tester, and R. Munns. 2007. HKT1; 5-like cation transporters linked to Na+ exclusion loci in wheat, Nax2 and Kna1. Plant Physiol. 143(4): 1918–1928.

Caicedo, A.L., S.H. Williamson, R.D. Hernandez, A. Boyko, A. Fledel-Alon, T.L. York, N.R. Polato, K.M. Olsen, R. Nielsen, S.R. McCouch, and others. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet 3(9): e163.

Cheng, X., R.K. Sardana, and I. Altosaar. 1998. Rice transformation by Agrobacterium infection. Recomb. Proteins from Plants Prod. Isol. Clin. Useful Compd.: 1–9.

Chinnusamy, V., a Jagendorf, and J.K. Zhu. 2005. Understanding and improving salt tolerance in plants. Crop Sci. 45: 437–448.

Cotsaftis, O., D. Plett, N. Shirley, M. Tester, and M. Hrmova. 2012. A two-staged model of Na+ exclusion in rice explained by 3D modeling of HKT transporters and alternative splicing. PLoS One 7(7): e39865.

Cui, H., Y. Takeoka, and T. Wada. 1995. Effect of Sodium Chloride on the Panicle and Spikelet Morphogenesis in Rice II. Developmental morphology of the panicle. Japanese J. Crop Sci. 64: 593.

Curtis, M.D., and U. Grossniklaus. 2003. A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. 133(2): 462–469.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, and others. 2011. The variant call format and VCFtools. Bioinformatics 27(15): 2156–2158.

Davenport, R.J., A. Muñoz-Mayor, D. Jha, P.A. Essah, A. Rus, and M. Tester. 2007. The Na+ transporter AtHKT1;1 controls retrieval of Na + from the xylem in Arabidopsis. Plant, Cell Environ. 30(2007): 497–507.

Eizenga, G.C., M. Ali, R.J. Bryant, K.M. Yeater, A.M. McClung, S.R. McCouch, and others. 2014. Registration of the rice diversity panel 1 for genomewide association studies. J. Plant Regist. 8(1): 109–116.

El-Hendawy, S.E., Y. Hu, and U. Schmidhalter. 2005. Growth, ion content, gas exchange, and water relations of wheat genotypes differing in salt tolerances. Crop Pasture Sci. 56(2): 123–134.

Fageria, N.K. 1985. Salt tolerance of rice cultivars. Plant Soil 88(2): 237–243.

Famoso, A.N., K. Zhao, R.T. Clark, C.-W. Tung, M.H. Wright, C. Bustamante, L. V Kochian, and S.R. McCouch. 2011. Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7(8): e1002221.

Flowers, T.J. 2004. Improving crop salt tolerance. J. Exp. Bot. 55(396): 307–319.

Flowers, T.J., and A.R. Yeo. 1995. Breeding for salinity resistance in crop plants: where next? Funct. Plant Biol. 22(6): 875–884.

Genc, Y., G.K. Mcdonald, and M. Tester. 2007. Reassessment of tissue Na+ concentration as a criterion for salinity tolerance in bread wheat. Plant. Cell Environ. 30(11): 1486–1498.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, R. Thompson, D. Butler, and others. 2009. ASReml user guide release 3.0. VSN Int. Ltd, Hemel Hempstead, UK.

Gregorio, G.B. 1997. Tagging salinity tolerance genes in rice using amplified fragment length polymorphism (AFLP).

Gregorio, G.B., D. Senadhira, R.D. Mendoza, N.L. Manigbas, J.P. Roxas, and C.Q. Guerta. 2002. Progress in breeding for salinity tolerance and associated abiotic stresses in rice. F. Crop. Res. 76(2): 91–101.

Heenan, D.P., L.G. Lewin, and D.W. McCaffery. 1988. Salinity tolerance in rice varieties at different growth stages. Aust. J. Exp. Agric. 28(3): 343–349.

Horie, T., F. Hauser, and J.I. Schroeder. 2009. HKT transporter-mediated salinity resistance mechanisms in Arabidopsis and monocot crop plants. Trends Plant Sci. 14(12): 660–668.

Huang, X., N. Kurata, X. Wei, Z. Wang, A. Wang, Q. Zhao, Y. Zhao, K. Liu, G. Dong, Q. Zhan, C. Li, A. Fujiyama, A. Toyoda, T. Lu, Q. Feng, Q. Qian, and J. Li. 2012. A map of rice genome variation reveals the origin of cultivated rice. Nature 490(7421): 497–501.

Huang, S., W. Spielmeyer, E.S. Lagudah, R.A. James, J.D. Platten, E.S. Dennis, and R. Munns. 2006. A sodium transporter (HKT7) is a candidate for Nax1, a gene for salt tolerance in durum wheat. Plant Physiol. 142(4): 1718–1727.

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, M. Li, D. Fan, Y. Guo, A. Wang, L. Wang, L. Deng, W. Li, Y. Lu, Q. Weng, K. Liu, T. Huang, T. Zhou, Y. Jing, W. Li, Z. Lin, E.S. Buckler, Q. Qian, Q. Zhang, J. Li, and B. Han. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Publ. Gr. 42(11): 961–967Available at http://dx.doi.org/10.1038/ng.695.

Ismail, A.M., M.J. Thomson, R.K. Singh, G.B. Gregorio, and D.J. Mackill. 2008. Designing rice varieties adapted to coastal areas of South and Southeast Asia. J. Indian Soc. Coast. Agric. Res. 26: 69–73.

Ismail, A.M., M.J. Thomson, G. V Vergara, M.A. Rahman, R.K. Singh, G.B. Gregorio, and D.J. Mackill. 2010. Designing Resilient Rice Varieties for Coastal Deltas Using Modern

Breeding Tools. Trop. Deltas Coast. Zo.: 154.

Ismail, A., and T. Tuong. 2009. Brackish water coastal zones of the monsoon tropics: challenges and opportunities. Nat. Resour. Manag. poverty Reduct. Environ. Sustain. rice-based Syst.: 113–121.

Jabnoune, M., S. Espeout, D. Mieulet, C. Fizames, J.-L. Verdeil, G. Conéjéro, A. Rodríguez-Navarro, H. Sentenac, E. Guiderdoni, C. Abdelly, and A.-A. Véry. 2009. Diversity in expression patterns and functional properties in the rice HKT transporter family. Plant Physiol. 150(August): 1955–1971.

Kang, H.M., N. a Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178(3): 1709–23.

Kent, W.J. 2002. BLAT—the BLAST-like alignment tool. Genome Res. 12(4): 656–664.

Khan, M.S.A., A. Hamid, and M.A. Karim. 1997. Effect of sodium chloride on germination and seedling characters of different types of rice (Oryza sativa L.). J. Agron. Crop Sci. 179(3): 163–169.

Khatun, S., and T.J. Flowers. 1995. Effects of salinity on seed set in rice. Plant. Cell Environ. 18(1): 61–67.

Koyama, M.L., A. Levesley, R.M.D. Koebner, T.J. Flowers, and A.R. Yeo. 2001. Quantitative Trait Loci for Component Physiological Traits Determining Salt Tolerance in Rice 1. 125(January): 406–422.

Laurie, S., K.A. Feeney, F.J.M. Maathuis, P.J. Heard, S.J. Brown, and R.A. Leigh. 2002. A role for HKT1 in sodium uptake by wheat roots. Plant J. 32(2): 139–149.

Lee, K.-S., W.-Y. Choi, J.-C. Ko, T.-S. Kim, and G.B. Gregorio. 2003. Salinity tolerance of japonica and indica rice (Oryza sativa L.) at the seedling stage. Planta 216(6): 1043–1046.

Leigh, J.W., and D. Bryant. 2015. popart: full-feature software for haplotype network construction. Methods Ecol. Evol. 6(9): 1110–1116.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16): 2078–2079.

Li, J., and L. Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). 95(3): 221–227.

Lin, H.X., M.Z. Zhu, M. Yano, J.P. Gao, Z.W. Liang, W. a Su, X.H. Hu, Z.H. Ren, and D.Y. Chao. 2004. QTLs for Na+ and K+ uptake of the shoots and roots controlling rice salt tolerance. Theor. Appl. Genet. 108(2): 253–60.

Livak, K.J., and T.D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25(4): 402–8.

Lutts, S., J.M. Kinet, and J. Bouharmont. 1995. Changes in plant response to NaCl during

development of rice (Oryza sativa L.) varieties differing in salinity resistance. J. Exp. Bot. 46(12): 1843–1852.

Mäser, P., B. Eckelman, R. Vaidyanathan, T. Horie, D.J. Fairbairn, M. Kubo, M. Yamagami, K. Yamaguchi, M. Nishimura, N. Uozumi, W. Robertson, M.R. Sussman, and J.I. Schroeder. 2002. Altered shoot/root Na+ distribution and bifurcating salt sensitivity in Arabidopsis by genetic disruption of the Na+ transporter AtHKT1. FEBS Lett. 531(2): 157–161.

Mather, K.A., A.L. Caicedo, N.R. Polato, K.M. Olsen, S. McCouch, and M.D. Purugganan. 2007. The extent of linkage disequilibrium in rice (Oryza sativa L.). Genetics 177(4): 2223–2232.

McCouch, S.R., M.H. Wright, C.-W. Tung, L.G. Maron, K.L. McNally, M. Fitzgerald, N. Singh, G. DeClerck, F. Agosto-Perez, P. Korniliev, and others. 2016. Open access resources for genome-wide association mapping in rice. Nat. Commun. 7.

Mian, A., R.J.F.J. Oomen, S. Isayenkov, H. Sentenac, F.J.M. Maathuis, and A.-A. Véry. 2011. Over-expression of an Na+-and K+-permeable HKT transporter in barley improves salt tolerance. Plant J. 68(3): 468–479.

Miki, D., R. Itoh, and K. Shimamoto. 2005. RNA silencing of single and multiple members in a gene family of rice. Plant Physiol. 138(4): 1903–1913.

Miki, D., and K. Shimamoto. 2004. Simple RNAi vectors for stable and transient suppression of gene function in rice. Plant Cell Physiol. 45(4): 490–495.

Munns, R., R. a James, B. Xu, A. Athman, S.J. Conn, C. Jordans, C.S. Byrt, R. a Hare, S.D. Tyerman, M. Tester, D. Plett, and M. Gilliham. 2012. Wheat grain yield on saline soils is improved by an ancestral Na^+ transporter gene. Nat. Biotechnol. 30(4): 360–4.

Munns, R., and M. Tester. 2008. Mechanisms of salinity tolerance. Annu. Rev. Plant Biol. 59: 651–81.

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich, and E.S. Buckler. 2009. Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21(August): 2194–2202.

Nakagawa, T., T. Kurose, T. Hino, K. Tanaka, M. Kawamukai, Y. Niwa, K. Toyooka, K. Matsuoka, T. Jinbo, and T. Kimura. 2007. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J. Biosci. Bioeng. 104(1): 34–41Available at http://www.sciencedirect.com/science/article/pii/S1389172307701160.

Negrão, S., M. Cecília Almadanim, I.S. Pires, I.A. Abreu, J. Maroco, B. Courtois, G.B. Gregorio, K.L. McNally, and M. Margarida Oliveira. 2013. New allelic variants found in key rice salt-tolerance genes: an association study. Plant Biotechnol. J. 11(1): 87–100.

Negrão, S., B. Courtois, N. Ahmadi, I. Abreu, N. Saibo, and M.M. Oliveira. 2011. Recent updates on salinity stress in rice: from physiological to molecular responses. CRC. Crit. Rev. Plant Sci. 30(4): 329–377.

Platten, J.D., J. a Egdane, and A.M. Ismail. 2013. Salinity tolerance, Na+ exclusion and allele mining of HKT1;5 in Oryza sativa and O. glaberrima: many sources, many genes, one mechanism? BMC Plant Biol. 13(1): 32.

Ren, Z.-H., J.-P. Gao, L.-G. Li, X.-L. Cai, W. Huang, D.-Y. Chao, M.-Z. Zhu, Z.-Y. Wang, S. Luan, and H.-X. Lin. 2005. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat. Genet. 37(10): 1141–6.

Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1): 139–140.

Rus, A., S. Yokoi, A. Sharkhuu, M. Reddy, B.H. Lee, T.K. Matsumoto, H. Koiwa, J.K. Zhu, R. a Bressan, and P.M. Hasegawa. 2001. AtHKT1 is a salt tolerance determinant that controls Na(+) entry into plant roots. Proc. Natl. Acad. Sci. U. S. A. 98: 14150–14155.

Senadhira, D. 1994. Rice and problem soils in South and Southeast Asia. IRRI Disc. Pap. Ser. 4: 1–186.

Šidák, Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. J. Am. Stat. Assoc. 62(318): 626–633.

Sunarpi, T. Horie, J. Motoda, M. Kubo, H. Yang, K. Yoda, R. Horie, W.-Y. Chan, H.-Y. Leung, K. Hattori, M. Konomi, M. Osumi, M. Yamagami, J.I. Schroeder, and N. Uozumi. 2005. Enhanced salt tolerance mediated by AtHKT1 transporter-induced Na unloading from xylem vessels to xylem parenchyma cells. Plant J. 44(6): 928–38.

Suzuki, K., N. Yamaji, A. Costa, E. Okuma, N.I. Kobayashi, T. Kashiwagi, M. Katsuhara, C. Wang, K. Tanoi, Y. Murata, and others. 2016. OsHKT1; 4-mediated Na+ transport in stems contributes to Na+ exclusion from leaf blades of rice at the reproductive growth stage upon salt stress. BMC Plant Biol. 16(1): 1.

Sweeney, M., and S. McCouch. 2007. The complex history of the domestication of rice. Ann. Bot. 100(5): 951–957.

Thomson, M.J., M. Ocampo, J. Egdane, M.A. Rahman, A.G. Sajise, D.L. Adorada, E. Tumimbang-Raiz, E. Blumwald, Z.I. Seraj, R.K. Singh, G.B. Gregorio, and A.M. Ismail. 2010. Characterizing the Saltol Quantitative Trait Locus for Salinity Tolerance in Rice. Rice 3(2–3): 148–160.

Trapnell, C., L. Pachter, and S.L. Salzberg. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9): 1105–11.

Walia, H., C. Wilson, P. Condamine, X. Liu, A.M. Ismail, L. Zeng, S.I. Wanamaker, J. Mandal, J. Xu, X. Cui, and T.J. Close. 2005. Comparative Transcriptional Profiling of Two Contrasting Rice Genotypes under Salinity Stress during the Vegetative Growth Stage. Plant Physiol. 139: 822–835.

Wang, R., W. Jing, L. Xiao, Y. Jin, L. Shen, and W. Zhang. 2015. The rice high-affinity potassium Transporter1; 1 is involved in salt tolerance and regulated by an MYB-type transcription factor. Plant Physiol. 168(3): 1076–1090.

Wisser, R.J., J.M. Kolkman, M.E. Patzoldt, J.B. Holland, J. Yu, M. Krakowsky, R.J. Nelson, and P.J. Balint-Kurti. 2011. Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a GST gene. Proc. Natl. Acad. Sci. 108(18): 7339–7344.

Yeo, A.R., M.E. Yeo, S.A. Flowers, and T.J. Flowers. 1990. Screening of rice (Oryza sativa L.) genotypes for physiological characters contributing to salinity resistance, and their relationship to overall performance. Theor. Appl. Genet. 79(3): 377–384.

Yoshida, S., D.A. Forno, and J.H. Cock. 1976. Laboratory Manual for Physiological Studies of Rice.

Yu, J., and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17: 155–160.

Zeng, L., M.C. Shannon, and C.M. Grieve. 2002. Evaluation of salt tolerance in rice genotypes by multiple agronomic parameters. : 235–245.

Zhao, K., C.W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2: 467Available at http://www.ncbi.nlm.nih.gov/pubmed/21915109%5Cnhttp://www.nature.com/ncomms/journal/v2/n9/pdf/ncomms1467.pdf.

Zhao, K., M. Wright, J. Kimball, G. Eizenga, A. McClung, M. Kovach, W. Tyagi, M.L. Ali, C.W. Tung, A. Reynolds, C.D. Bustamante, and S.R. McCouch. 2010. Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. PLoS One 5(5).

2.6 Figures

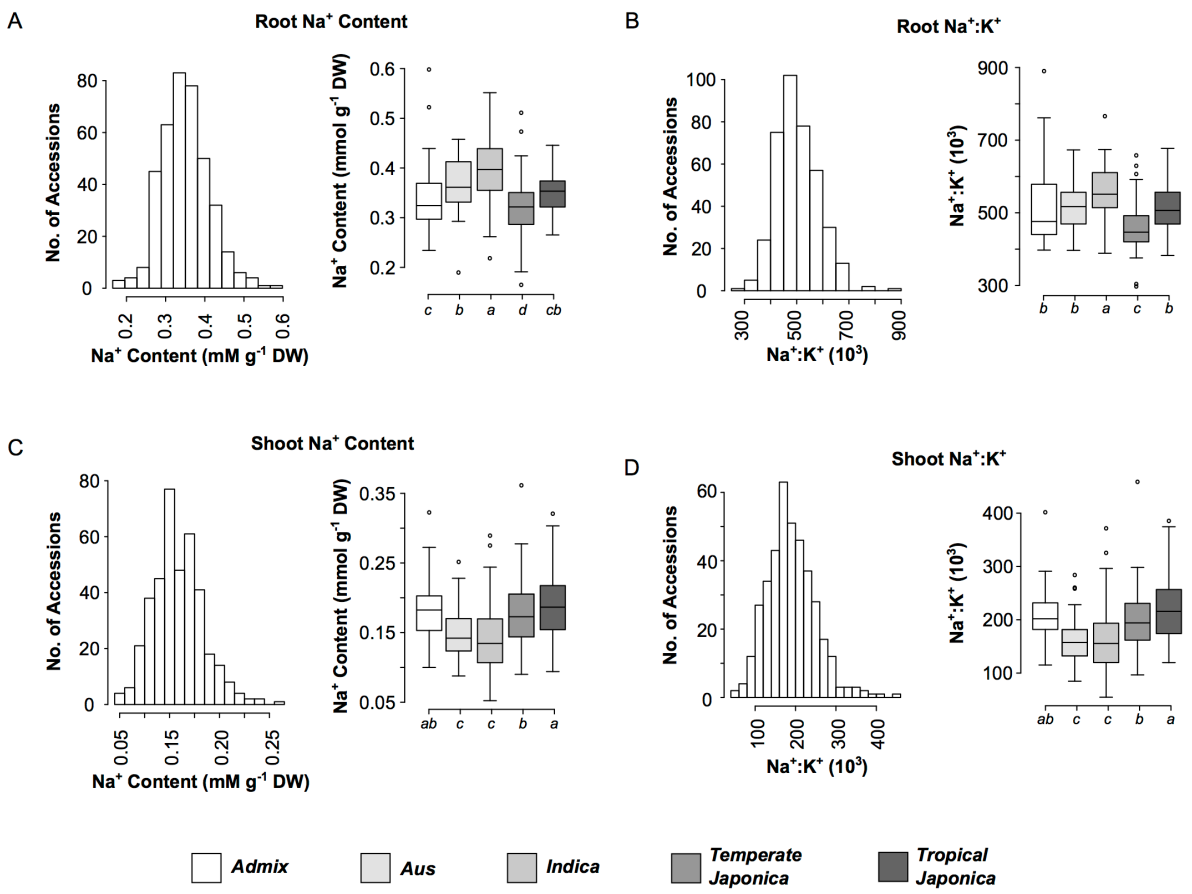


Fig 1. Phenotypic variation and distribution of root and shoot Na^+ content and Na^+:K^+ among the five subpopulations of RDP1. (A) Root Na^+ content, (B) root Na^+:K^+, (C) shoot Na^+ content, and (D) shoot Na^+:K^+. Accessions were assigned to each subpopulation according to Famoso *et al* (Famoso et al., 2011). Boxplots with the same letter indicate no significant difference as determined using ANOVA ($p < 0.05$). *Aromatic* accessions were excluded from the analysis due to low *n*.

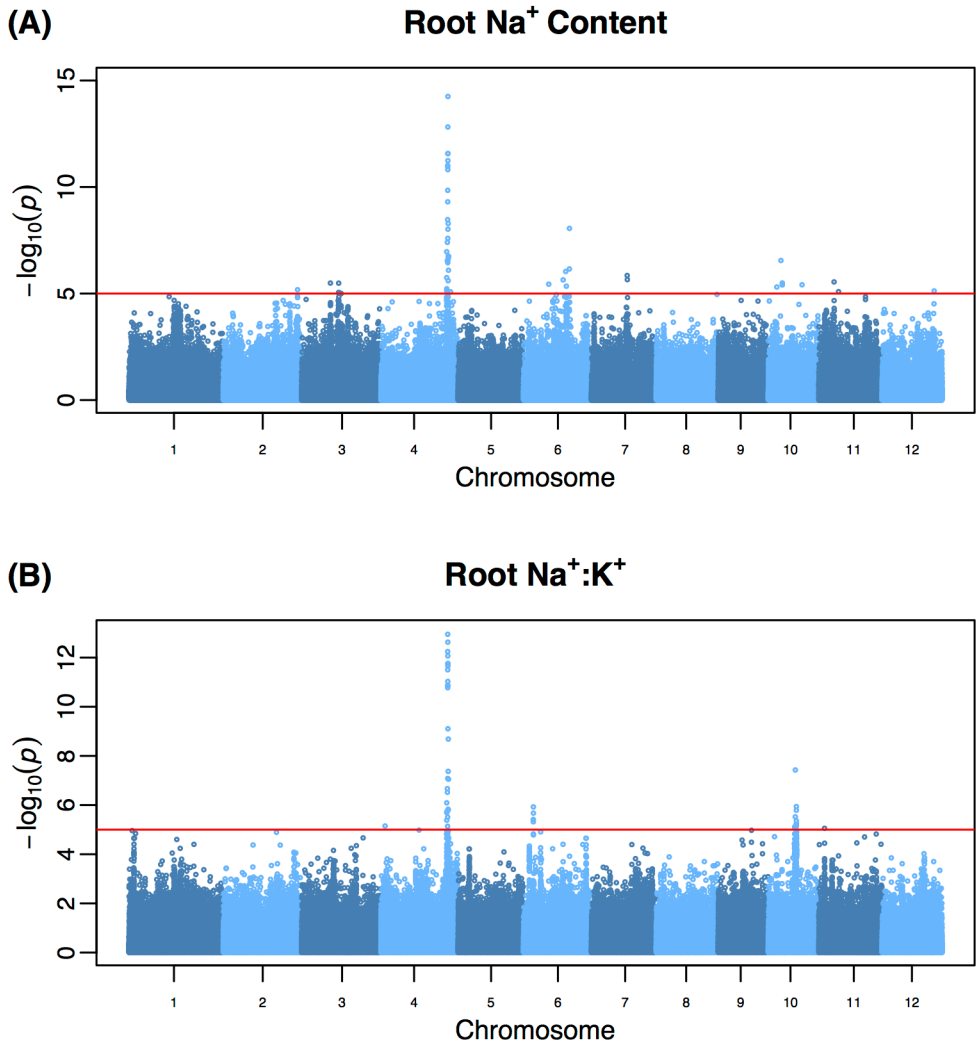


Fig 2. Genome-wide association analysis of Na^+ content and Na^+:K^+ in root tissue. (A) Root Na^+ content, (B) root Na^+:K^+. Genome-wide association (GWA) was performed using a mixed model that accounted for population structure and relatedness between accessions of RDP1 using 365 accessions of RDP1 and 397,812 SNPs. For each trait the least squares mean was used as the dependent variable The red horizontal line indicates a statistical significance threshold of $p < 10^{-5}$, and was determined using the M_{eff} method with an experiment-wise error rate of 0.05 (Li and Ji, 2005).

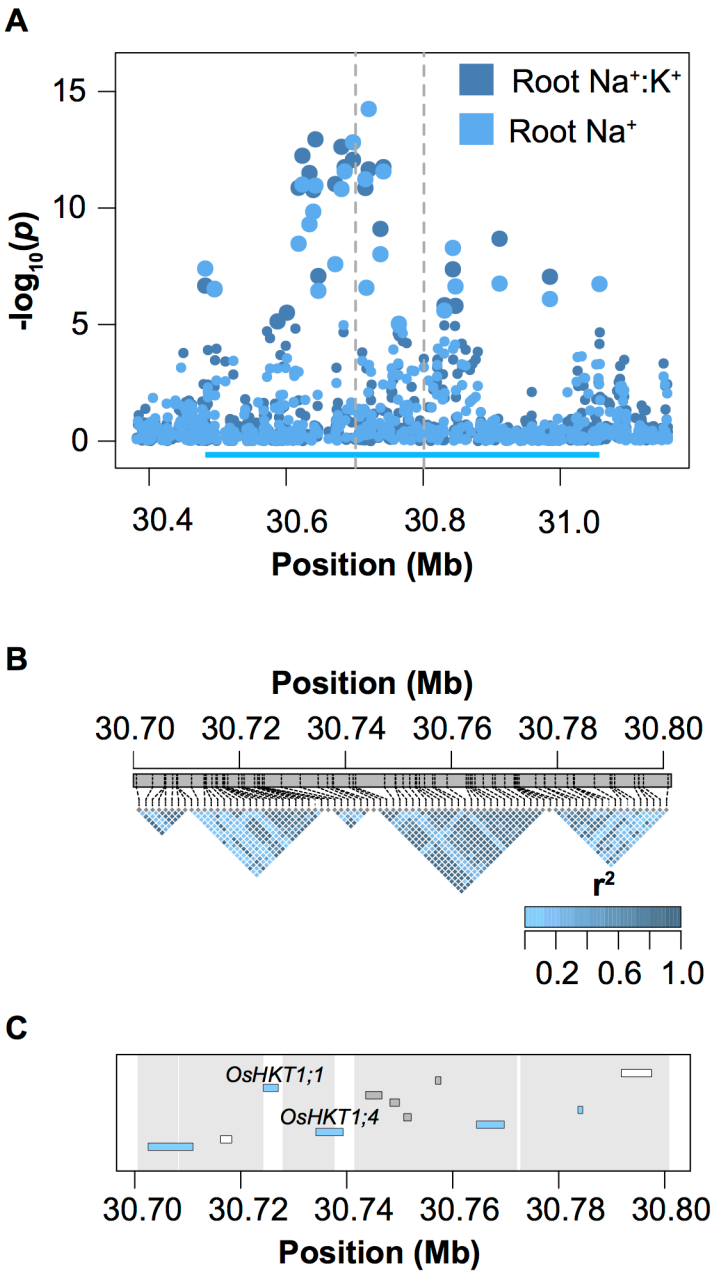


Fig 3. Genetic characterization of *RNC4*. (A) Regional manhattan plot summarizing GWA analysis of root Na^+ and root Na^+:K^+. The region defining *RNC4* is indicated with the cyan bar. (B) LD plots for a subset of five haplotype blocks within *RNC4*. The vertical broken gray lines in A indicate the region characterized by haplotype blocks. The genes present in this region are

illustrated in C. Genes encoding transposable elements are highlighted in the gene track in gray, while those encoding expressed proteins are highlighted in white. The regions defined by each block are indicated in gray.

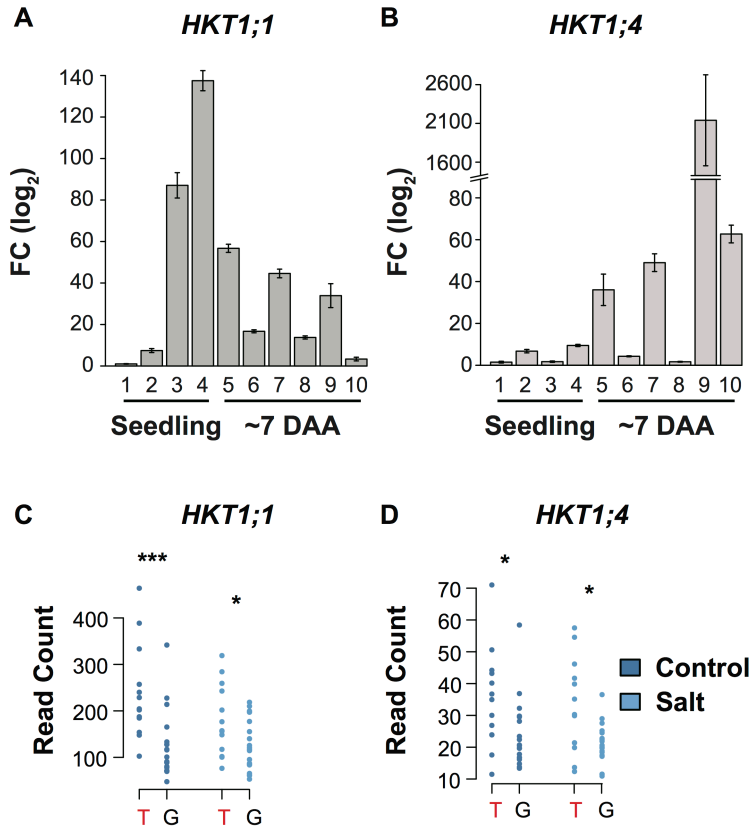


Fig 4. Characterization of *HKT1;1* and *HKT1;4* expression. Expression profiles of *HKT1;1* (A) and *HKT1;4* (B) in seeding and mature plants. Numbers below each of the bars indicate different tissues as follows: 1: seedling root, 2: seedling shoot, 3: blade of newest fully expanded leaf in seedling (leaf 3), 4: sheath of newest fully expanded leaf in seedling (leaf 3), 5: penultimate leaf sheath in mature plant, 6: penultimate leaf blade in mature plant, 7: flag leaf sheath in mature plant, 8: flag leaf blade in mature plant, 9: culm of mature plant and 10: panicle of mature plant. DAA: days after anthesis. (C, D) Dot plots comparing the expression of *HKT1;1* (C) and *HKT1;4* (D) between allelic groups at SNP-4-30535352 in control and saline conditions.

The minor allele genotype, which displays higher root Na^+ content, is indicated by red text. A Mann-Whitney test was performed within each treatment to determine differences between the two groups with asterisks indicating significance as determined using a one-way ANOVA: ***: $p < 0.001$; *: $p < 0.05$.

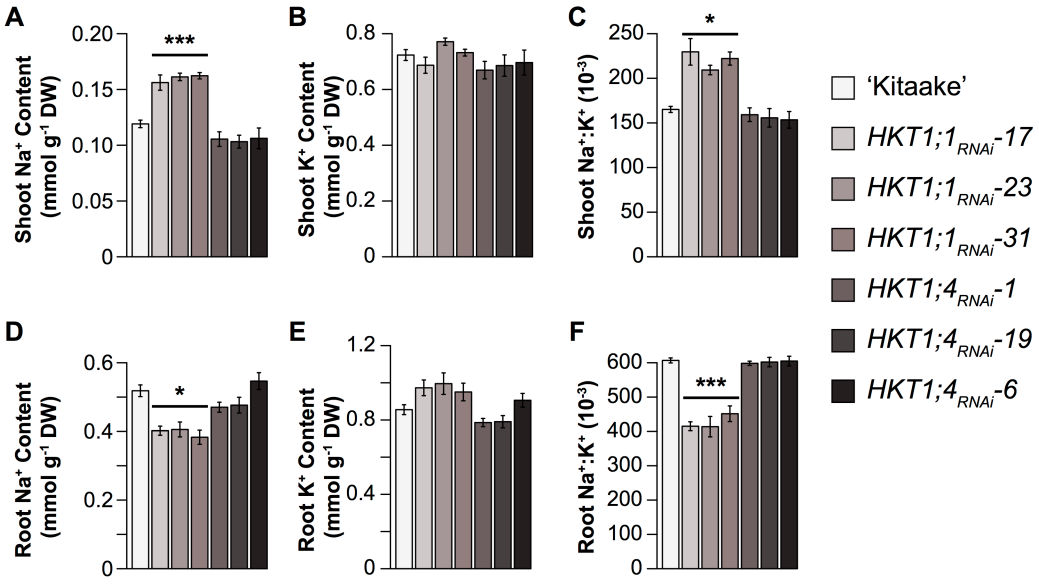


Fig 5. Root (A-C) and shoot (D-F) ion content of *HKT1;1* and *HKT1;4* RNAi plants.

Asterisks indicate significance as determined using Tukey's HSD: ***: $p < 0.001$; *: $p < 0.05$. Pairwise comparisons were made between lines within each treatment. Error bars represent standard error of the mean where n= 12-20 plants.

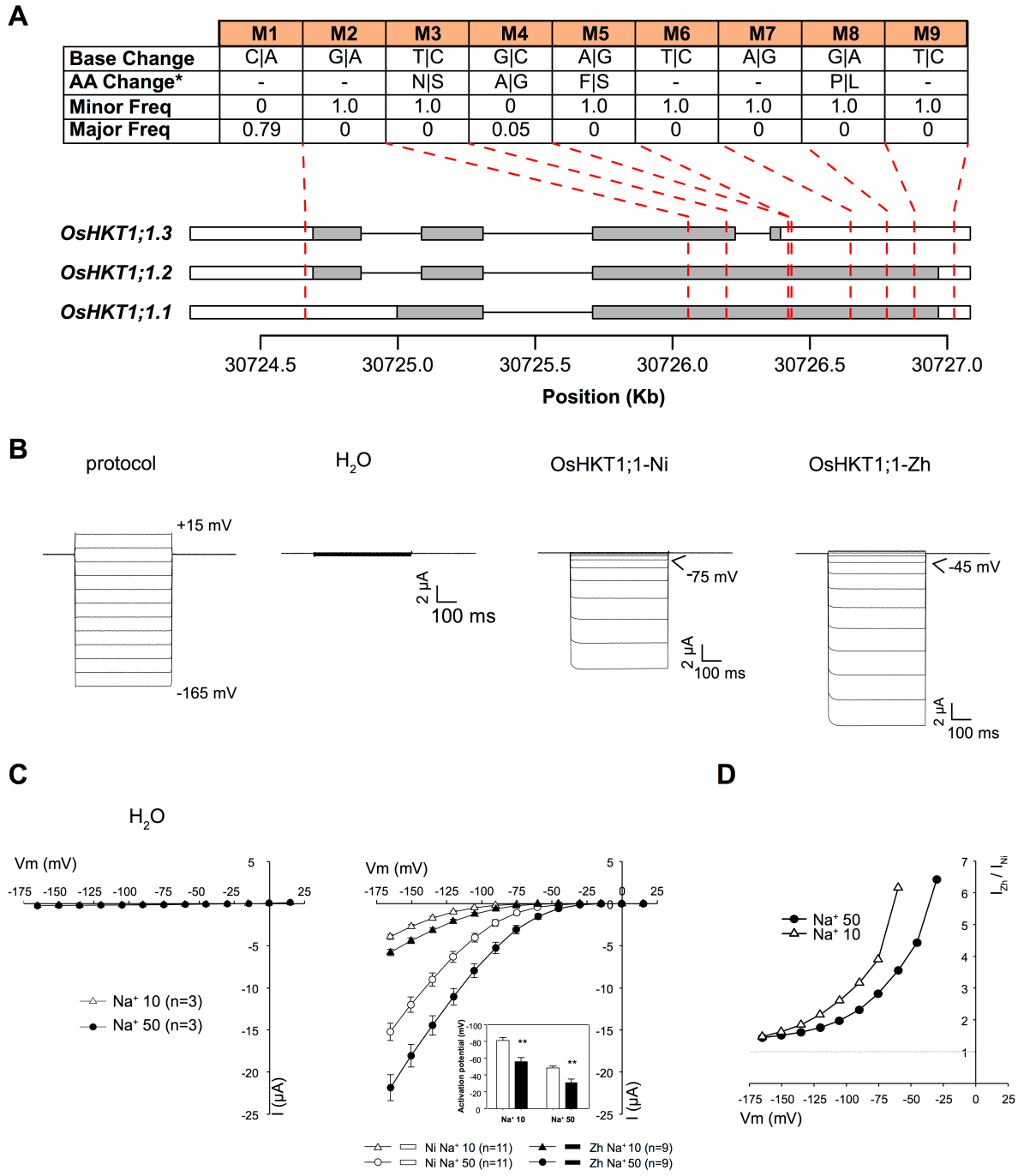


Fig 6. Characterization of high and low root Na^+ isoforms of HKT1;1 in *Xenopus* oocytes.

(A) Genetic variants within the ORF of *HKT1;1*. "Minor Freq" and "Major Freq" indicate the frequency of the alternate allele in the major (*n*=19) and minor (*n*=13) allelic groups at SNP-4-

30535352. “AA change” indicates the resulting changes in protein sequence. Synonymous mutations are indicated by “-”. The grey bars represent exons while the white bars represent the 3’ and 5’ UTRs. (B) Secondary structure of OsHKT1;1 polypeptide showing the position of AA changes, as exemplified between ‘Nipponbare’ and ‘Zhenshan 2’ variants. (C, D) Comparison of OsHKT1;1-Ni and –Zh targeting to the oocyte membrane by confocal imaging of GFP-tagged transporters. (C) Representative images of oocytes expressing Ni (top) or Zh (bottom) transporters. Emitted fluorescence was collected between 505 and 510 nm. Scale bar: 100 μm. (D) Comparison of fluorescence intensity spectra at the membrane in water-injected oocytes (control) and in oocytes expressing either of the HKT1;1 variants. Data are means ± SE. (E) Voltage-clamp protocol and corresponding representative current traces recorded in control oocytes or oocytes expressing the HKT1;1 variants, in 50 mM Na-glutamate-containing bath solution. (F) Current-voltage (I-V) relationships in control oocytes (left) and in HKT1;1-Ni or -Zh-expressing oocytes (right), in either 10 or 50 mM Na-glutamate-containing bath solutions. Data are means ± SE. Insert: Activation potential of HKT1;1-Ni or -Zh currents. Asterisks indicate significant difference in activation potential as determined using Student’s t test: **: $p < 0.005$. (G) HKT1;1-Zh to HKT1;1-Ni mean current ratio at varying membrane potentials, determined from I-V data shown in (F). Shown data in (E to G) were obtained in a single oocyte batch and are representative of three experiments performed in different oocyte batches.

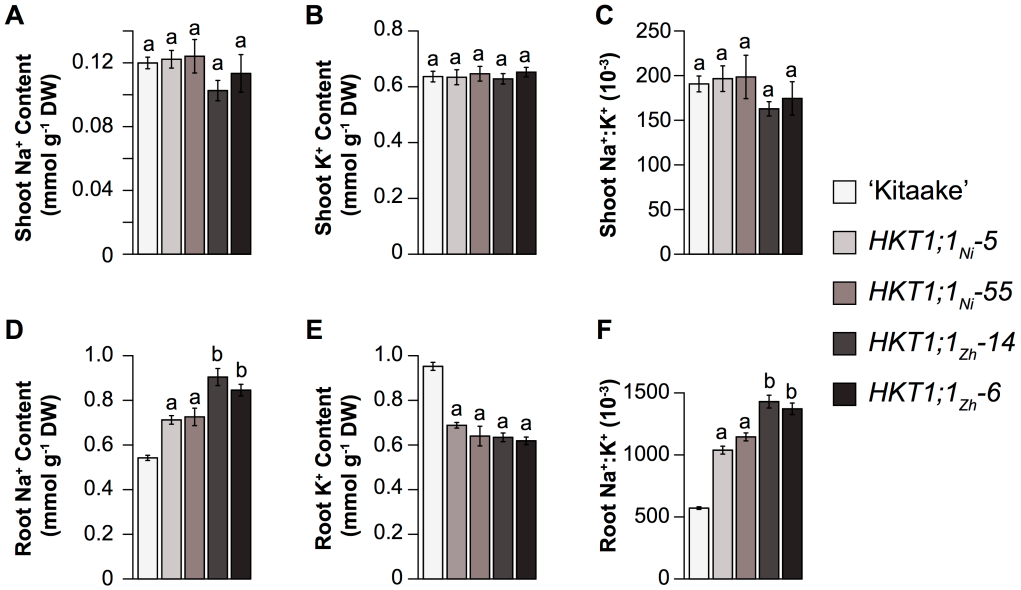


Fig 7. Shoot (A-C) and root (D-F) ion content for *HKT1;1* native overexpression lines. Statistical significance was determined using Tukey's HSD test between each line within treatments. Bars with the same letters indicate no significant difference ($p < 0.05$). Error bars represent standard error of the mean where n= 12-18 plants.

A

	H1	H2	H3	H4	H5	H6	H7	H8
Admix	0.26	0.51	0.14	0.08	0	0	0	0.01
Aro.	0.01	0.95	0.03	0	0	0	0.01	0
Aus	0.10	0.13	0.74	0	0.02	0	0	0.01
Ind. 1A	0.95	0.04	0	0	0.01	0	0	0.01
Ind. 1B	0.86	0.12	0.01	0.02	0.01	0	0	0
Ind. 2	0.76	0.16	0.02	0	0	0	0	0.06
Ind. 3	0.87	0.11	0	0	0	0.01	0	0.01
Indx	0.74	0.17	0.01	0.01	0.01	0.04	0	0.03
Temp.	0.01	0.06	0	0.67	0	0	0.27	0
Trop. 1	0.02	0.63	0.27	0.07	0	0	0.01	0
Trop. 2	0	0.76	0.24	0	0	0	0	0
Japx	0.03	0.44	0.04	0.48	0	0	0.03	0
Or-I	0.38	0.07	0.06	0	0.47	0.03	0	0
Or-II	0.67	0.04	0	0	0.27	0.03	0	0
Or-III	0.25	0.16	0	0.03	0.05	0.51	0.01	0

B

	H8	H7	H6	H5	H4	H3	H2	H1
H1	94.1	35.3	47.1	94.1	41.2	17.7	23.5	100
H2	17.7	76.5	76.5	29.4	70.6	94.1	100	
H3	11.8	70.6	70.6	23.5	64.7	100		
H4	35.3	82.4	94.1	47.1	100			
H5	88.2	41.2	52.9	100				
H6	41.2	88.2	100					
H7	29.4	100						
H8	100							

C

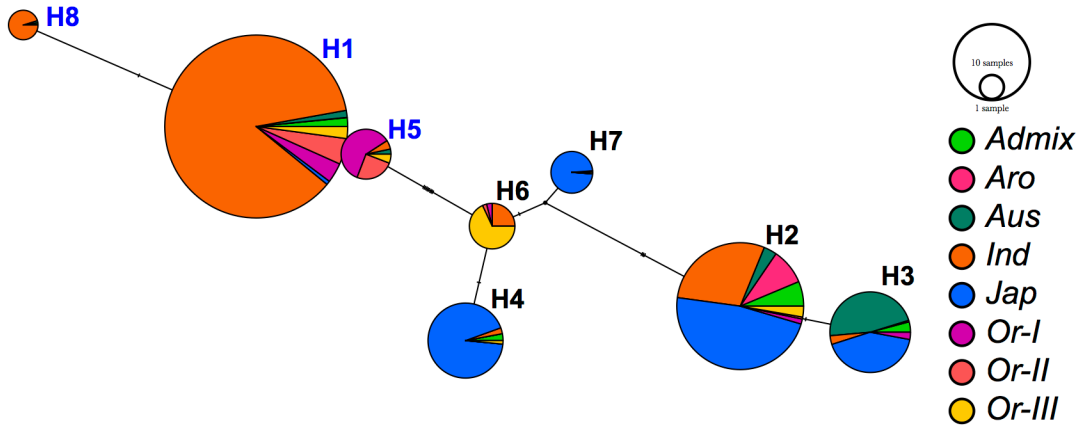


Fig 8. Haplotype analysis of a 37 kb region around HKT1;1 in cultivated and wild rice. (A) Haplotype frequencies in a panel of 3,024 cultivated rice accessions and 446 Oryza rufipogon

accessions. (B) Pairwise comparisons of sequence similarities in between the haplotypes in A. High root Na^+ haplotypes (i.e. those with the three non-synonymous SNPs are highlighted in blue) (C) Haplotype network for the 37 kb region surrounding *HKT1;1*. The size of each node is proportional to the number of accessions with the corresponding haplotype. Each node is separated into a pie chart, which indicates the number of accessions for each population with the haplotype. *Indica* (*Ind*) accessions are those belonging to *Ind 1A*, *Ind 1B*, *Ind 2*, *Ind 3*, and *Indx*. *Japonica* accessions are those belonging to *Temp*, *Trop 1*, *Trop 2*, and *Japx*.

2.7 Tables

Table I. Phenotypic and genetic correlation of root and shoot ion traits in RDP1. The lower diagonal represents the phenotypic Pearson correlation coefficients, while the upper diagonal represents the genetic correlation coefficients (n=383). Here, the genetic correlation between traits accounts for population structure and relatedness between accessions of RDP1.

Trait	Root Na^+	Root K^+	Root Na^+:K^+	Shoot Na^+	Shoot K^+	Shoot Na^+:K^+
Root Na^+	1.00	0.25	0.81	0.06	0.11	-0.07
Root K^+	0.47	1.00	-0.36	-0.21	0.09	-0.20
Root Na^+:K^+	0.66	-0.32	1.00	0.12	0.00	0.11
Shoot Na^+	0.00	-0.09	0.06	1.00	-0.17	0.96
Shoot K^+	0.14	0.14	0.04	-0.13	1.00	-0.41
Shoot Na^+:K^+	-0.02	-0.12	0.06	0.96	-0.35	1.00

Table II. Proportion of variation explained by a subset of the haplotype blocks within *RNC4*. Haplotype blocks were determined using the 4Gamete rule in Haploview with a recombination threshold of > 2%. Analysis of variance (ANOVA) was used to estimate proportion of phenotypic variance accounted for by each block after adjusting for population structure effects.

		Root Ion			Shoot Ion		
Block	**Position (bp)**	**Na^+**	**K^+**	**Na^+:K^+**	**Na^+**	**K^+**	**Na^+:K^+**
1	30,700,577 – 30,708,292	0.013	0.049	0.027	0.013	0.058	0.012
2	30,708,321 – 30,724,332	0.045	0.036	0.054	0.035	0.066	0.040
3	30,727,920 – 30,737,580	0.160	0.024	0.175	0.037	0.029	0.032
4	30,740,743 – 30,771,971	0.012	0.123	0.069	0.145	0.094	0.125
5	30,772,483 – 30,795,385	0.060	0.033	0.079	0.037	0.050	0.036

2.8 Supplemental Materials

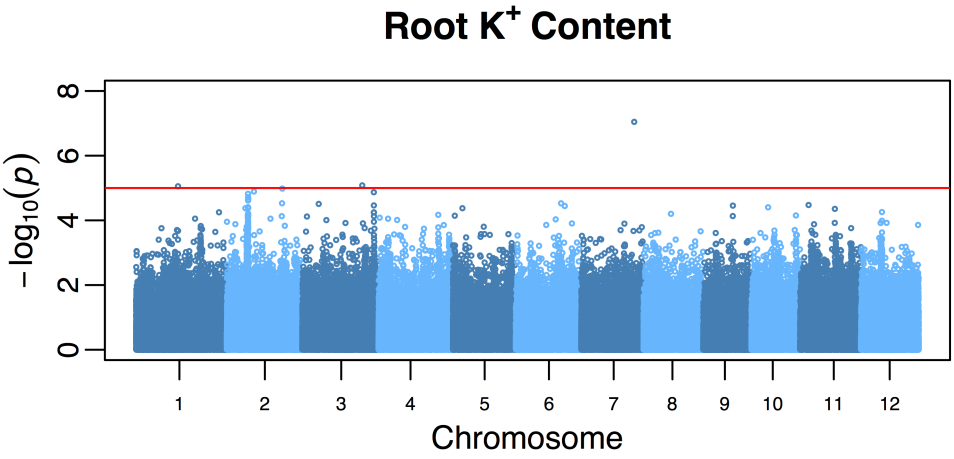


Figure S1. GWA of root K^+ content in salt conditions. The red horizontal line indicates the genome-wide p-value threshold ($p < 10^{-5}$).

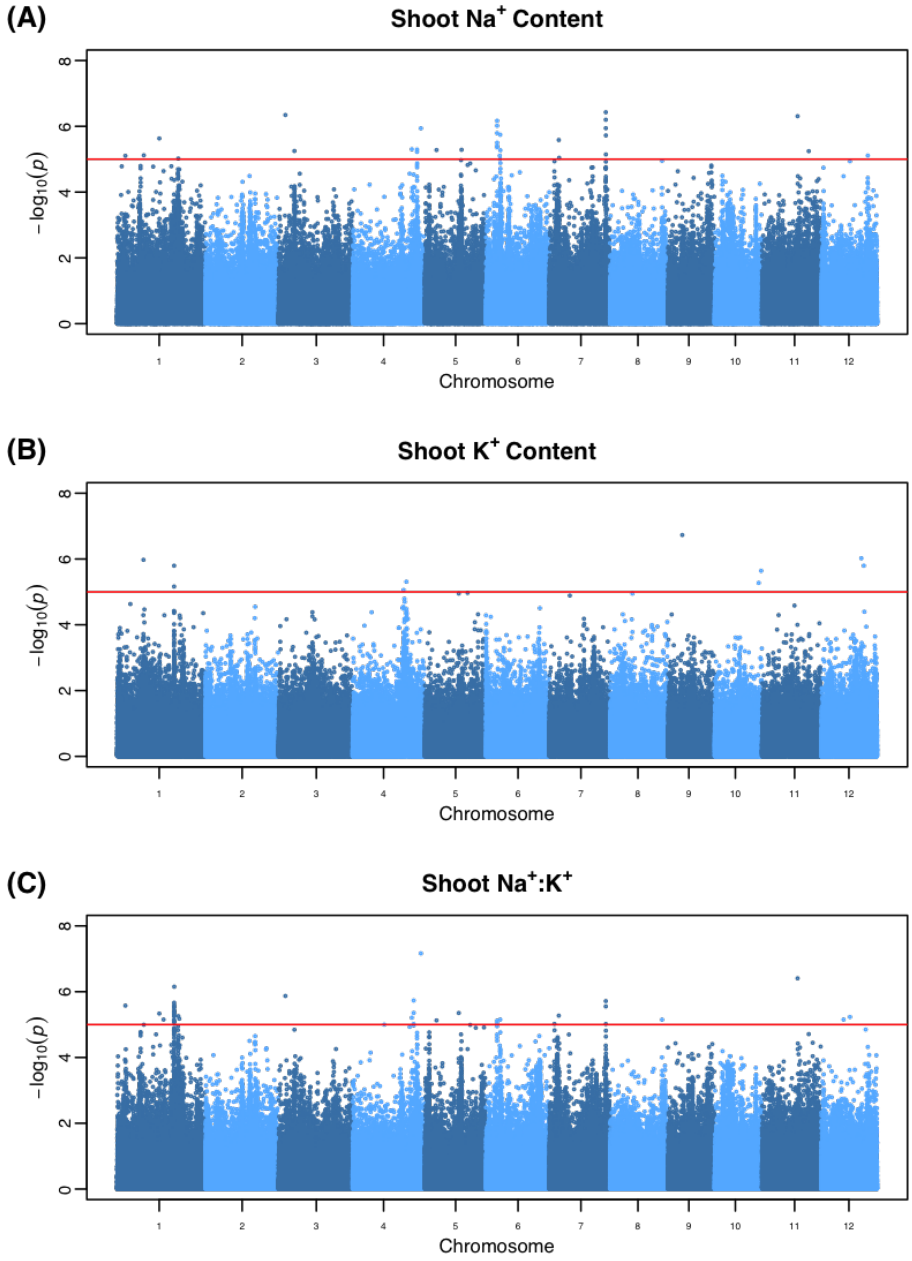


Figure S2. GWA of shoot ion traits in salt conditions. The red horizontal line indicates the genome-wide p-value threshold ($p < 10^{-5}$).

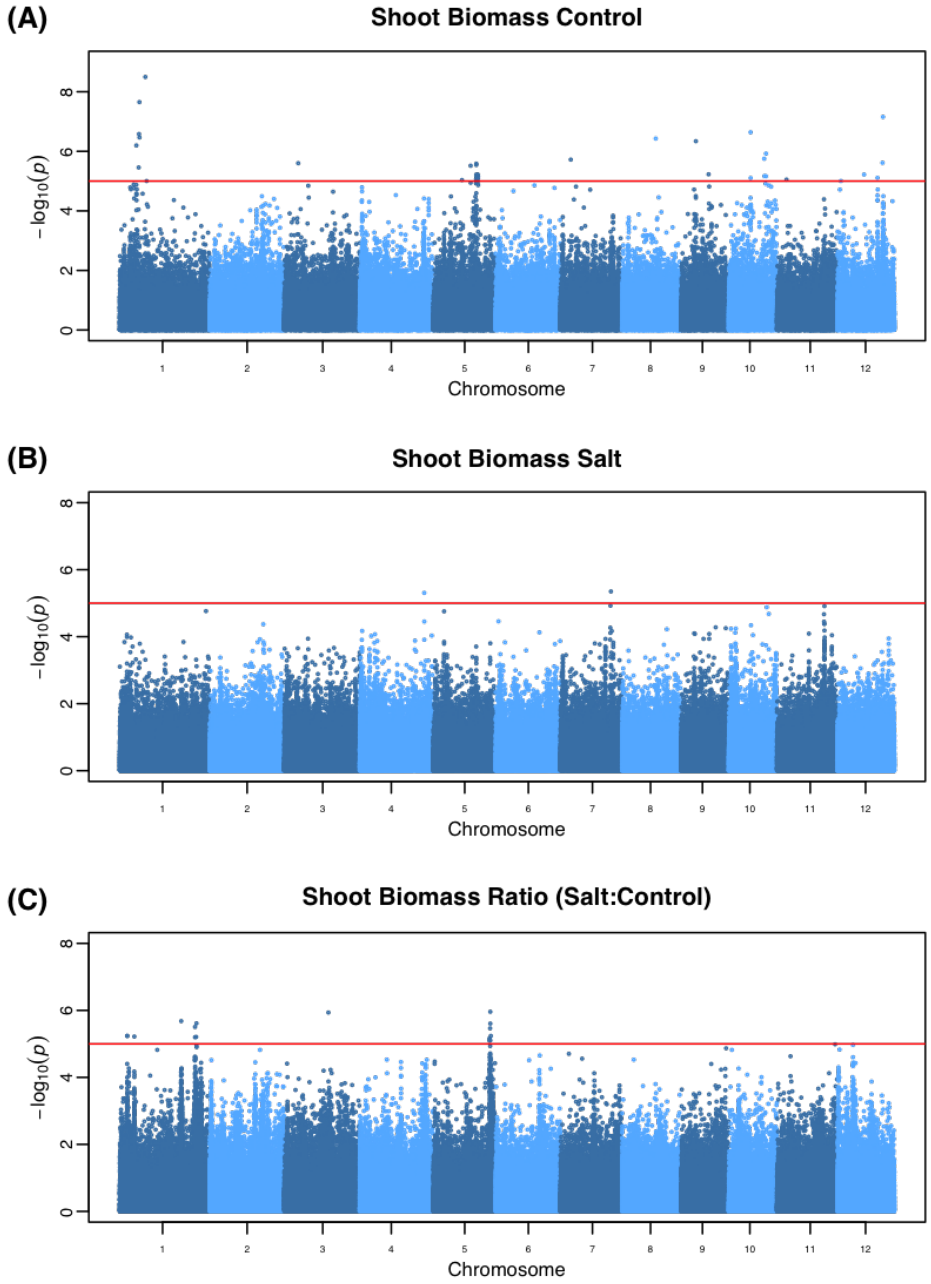


Figure S3. GWA of shoot biomass traits in salt and saline conditions. The ratio of biomass in salt to control was used to control for the inherent differences in growth rate among accessions. The red horizontal line indicates the genome-wide p-value threshold ($p < 10^{-5}$).

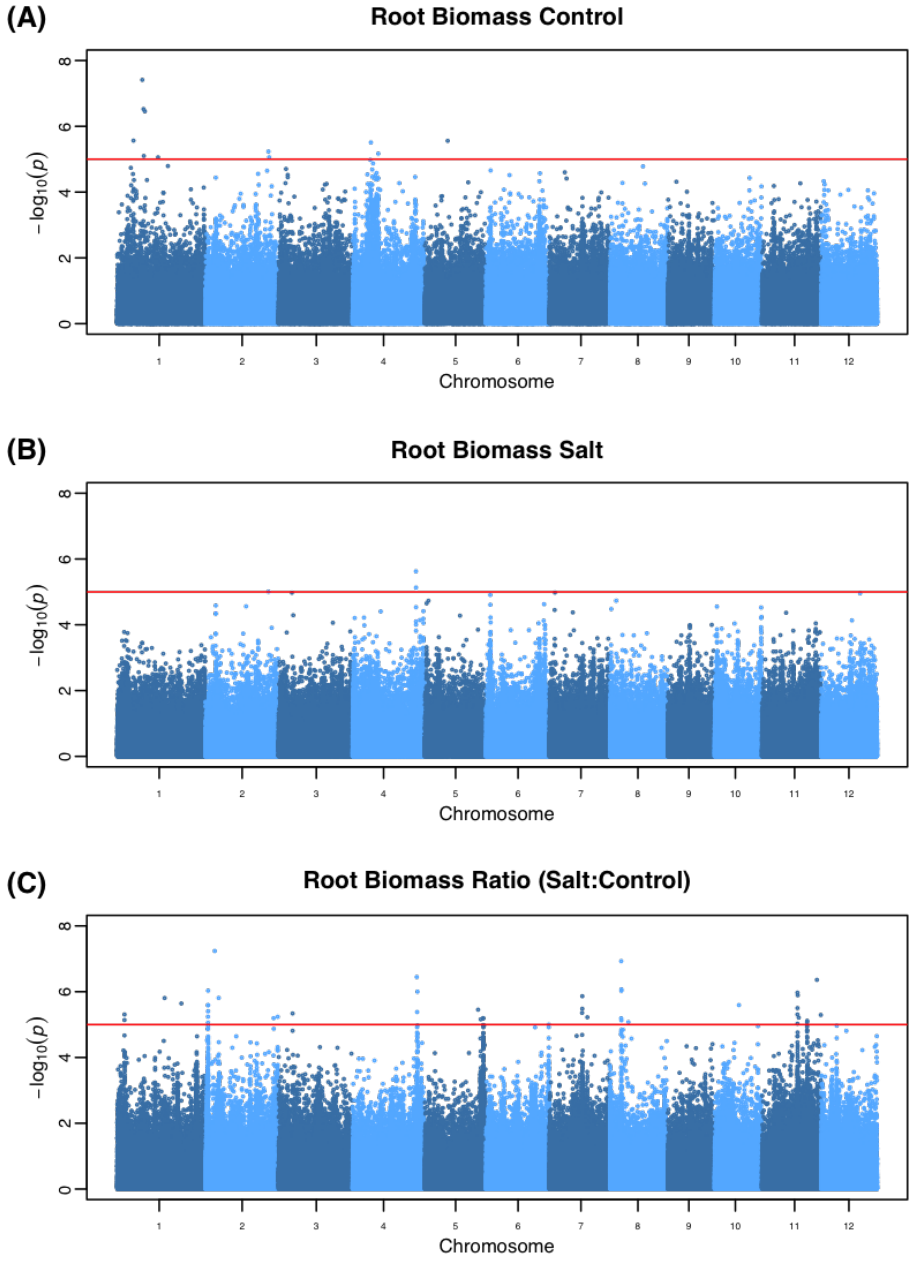


Figure S4. GWA of root biomass traits in salt and saline conditions. The ratio of biomass in salt to control was used to control for the inherent differences in growth rate among accessions. The red horizontal line indicates the genome-wide p-value threshold ($p < 10^{-5}$).

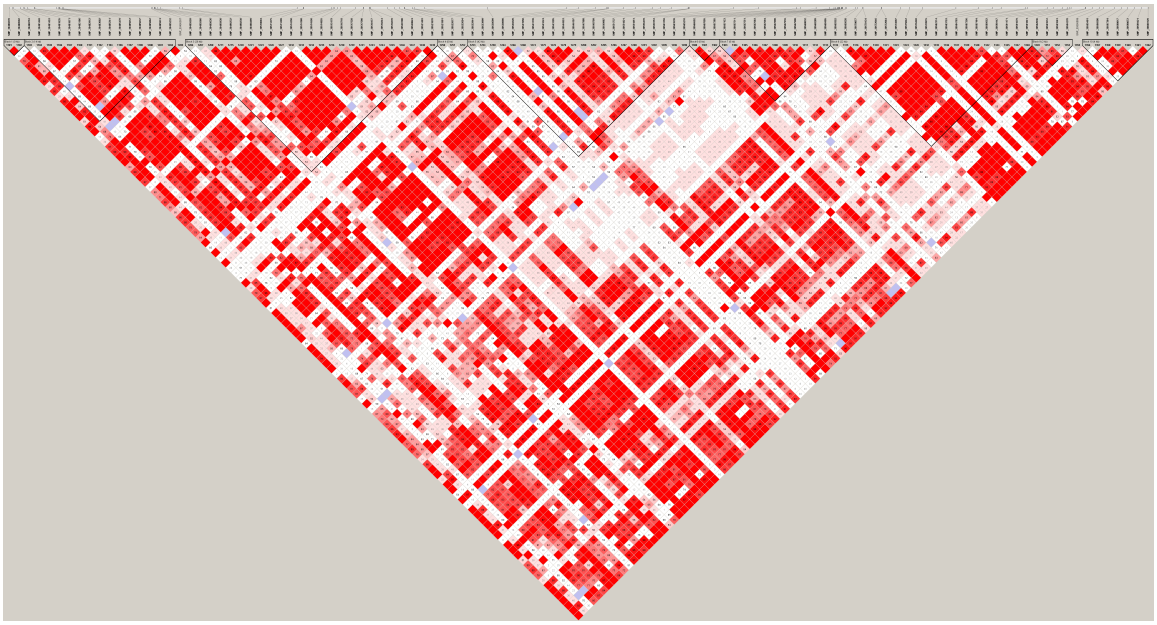


Figure S5. Haplotype Blocks for the region spanning 30.48-30.6 Mb on chromosome 4.

Figure S6. Haplotype Blocks for the region spanning 30.6-30.7 Mb on chromosome 4.

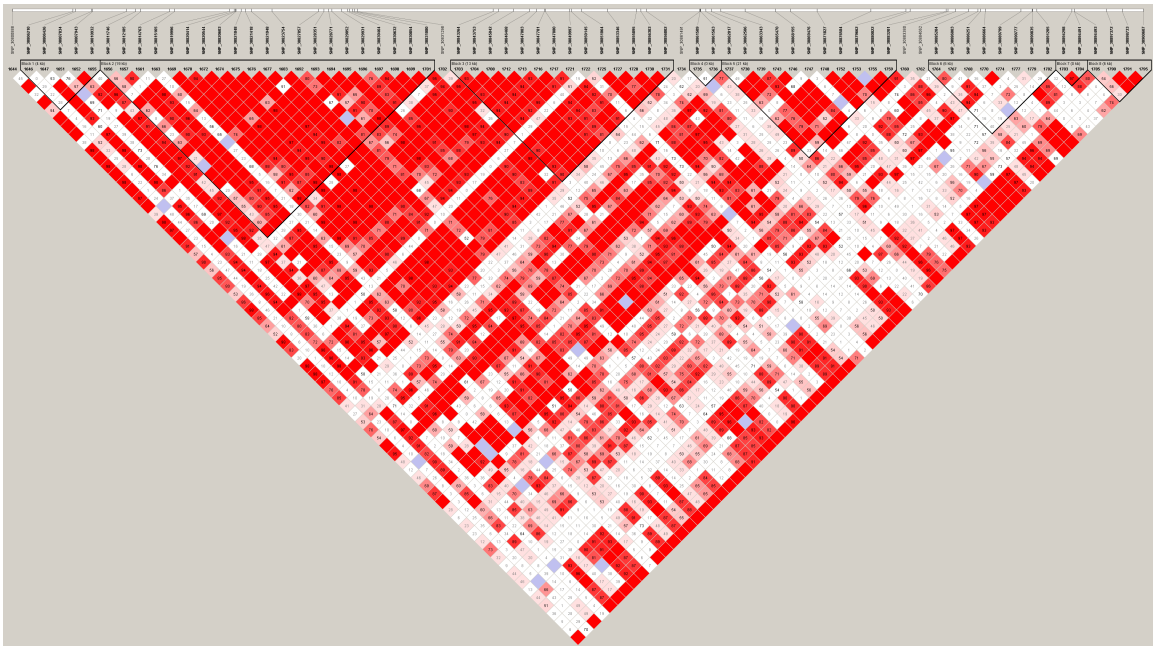


Figure S7. Haplotype Blocks for the region spanning 30.8-30.9 Mb on chromosome 4.

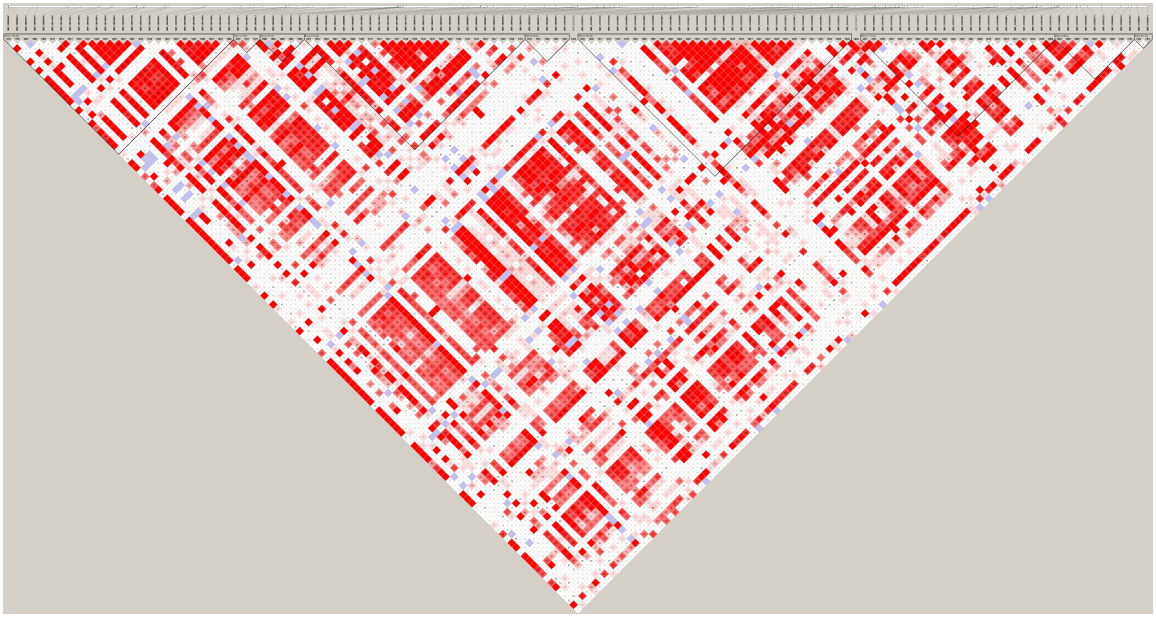


Figure S8. Haplotype Blocks for the region spanning 30.9-31.06 Mb on chromosome 4.

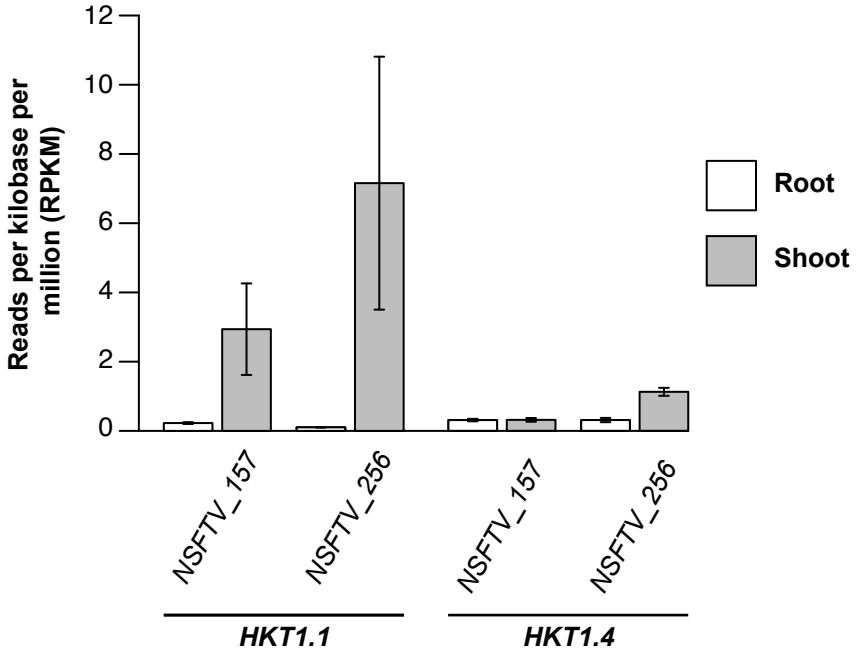


Fig S9. Expression of *HKT1;1* and *HKT1;4* in root and shoot tissue. RNA sequencing of root and shoot tissue was performed with two accessions of RDP1 at 10 days after transplanting (12 day old plants) when the first tiller was visibly emerging. The expression levels of HKT1;1 and HKT1;4 are expressed as reads per kilobase per million mapped reads (RPKM). RPKM was determined using EdgeR (Robinson et al., 2010) and MSUv7 annotation was used to determine the length of each gene.

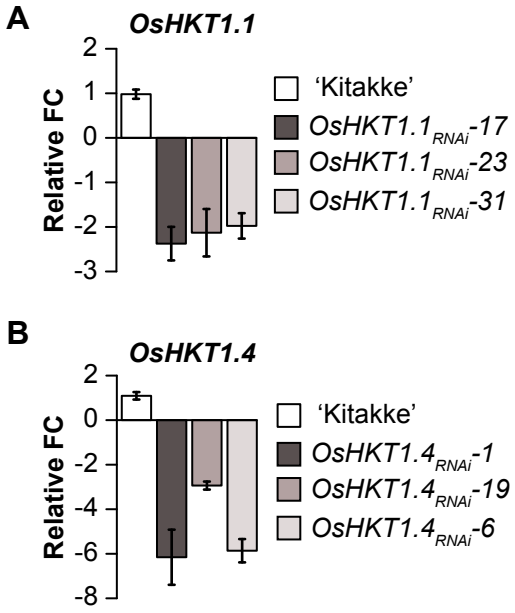


Figure S10. Expression of *HKT1;1* (A) and *HKT1;4* (B) in RNAi plants. For *HKT1;1* gene expression was quantified in whole shoot tissue in four-day-old plants, while for *HKT1;4* expression was quantified from flag leaf tissue at anthesis. All expression is relative to 'Kitaake'. Error bars represent standard error of the mean where n = 4 biological replicates.

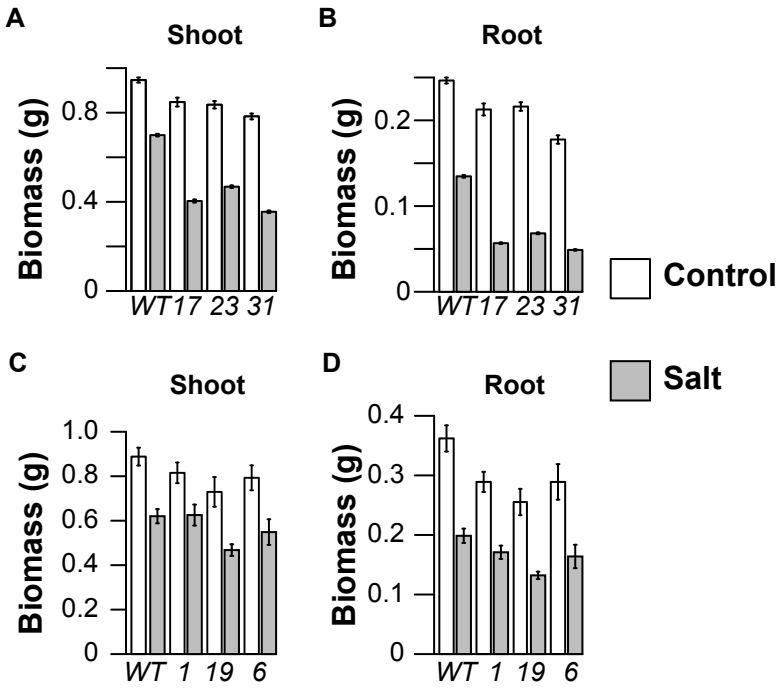


Figure S11. Shoot (A,C) and root (B,D) growth in RNAi lines. Error bars represent standard error of the mean where n = 12-20 plants.

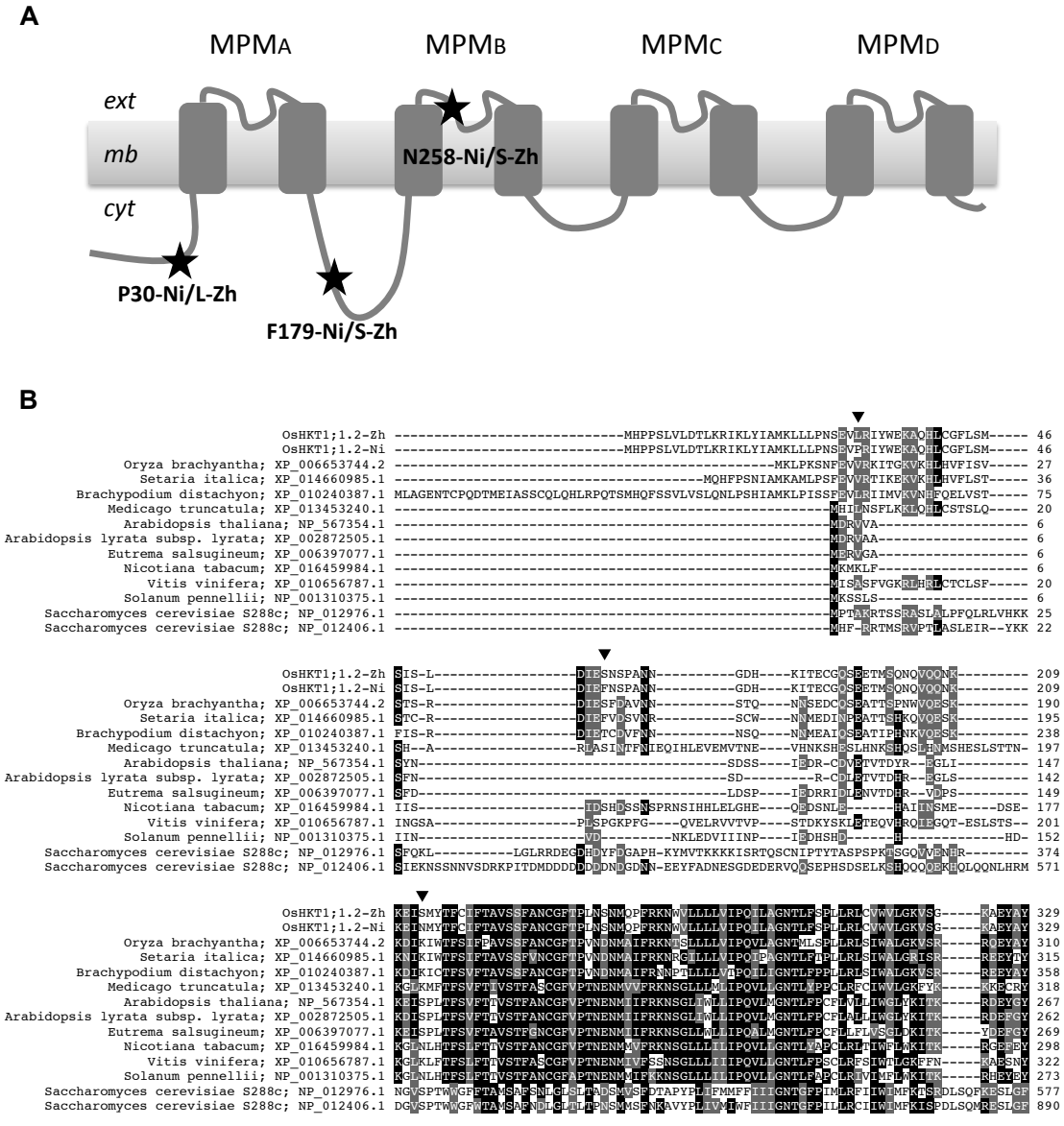


Figure S12. Position and conservation of non-synonymous variants in HKT1;1.2. (A) Secondary structure of OsHKT1;1 polypeptide indicating position of allelic variations between 'Nipponbare' (Ni) and 'Zhenshan 2' (Zh) isoforms. (B) Protein alignment of HKT1;1.2 and HKT1;1 orthologs from various species. Black triangles indicate position of three non-synonymous mutations present in HKT1;1-Zh.

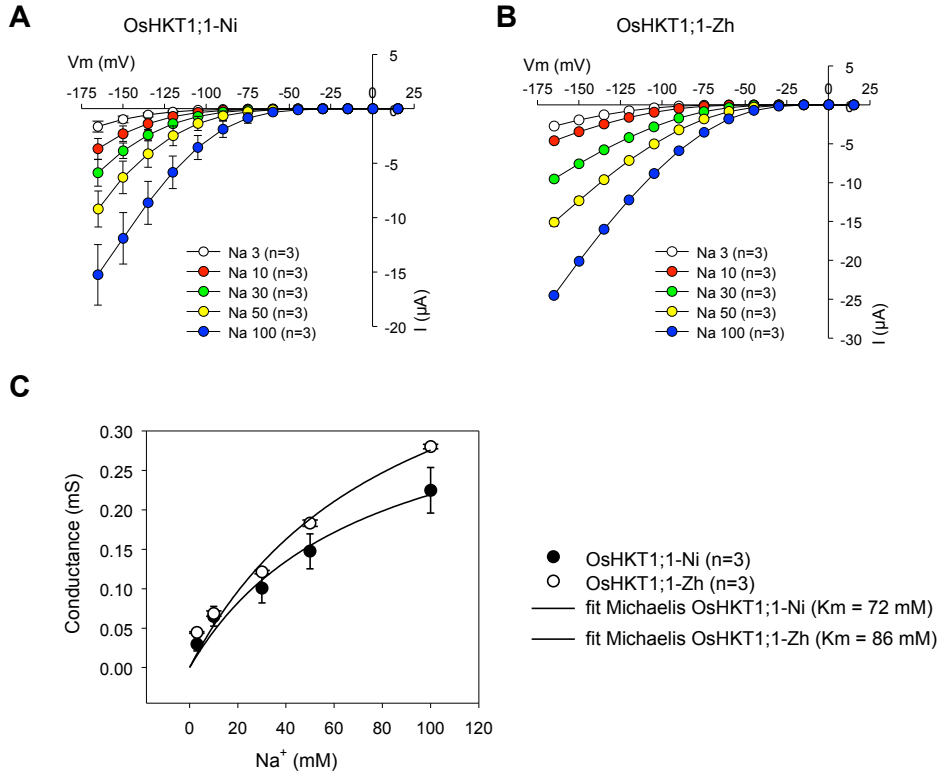


Fig S13. HKT1;1-Ni and HKT1;1-Zh display similar affinity for Na^+. (A, B) Current voltage (I-V) relationships determined in solutions containing 3 to 100 mM Na+-glutamate in HKT1;1-Ni (A) or HKT1;1-Zh (B) expressing oocytes from the same batch. (C) HKT1;1-Ni or -Zh inward conductance plotted versus external Na^+ concentration. Whole oocyte HKT1;1 inward conductances, were determined from I-V data shown in (A) and (B) at membrane voltages from -120 to -150 mV. HKT1;1-Ni and –Zh conductances were not significantly different as determined using Student's t test (p> 0.05). The concentration at which half saturation of the mean conductance occurred (apparent Km) and the maximal conductance were determined with hyperbolic fits (Michaelis-Menten equation): K_m = 72 mM and G_{max} = 0.38 mS for HKT1;1-Ni, K_m = 86 mM and Gmax = 0.51 mS for HKT1;1-Zh. Data in (A-C) are means ± SE.

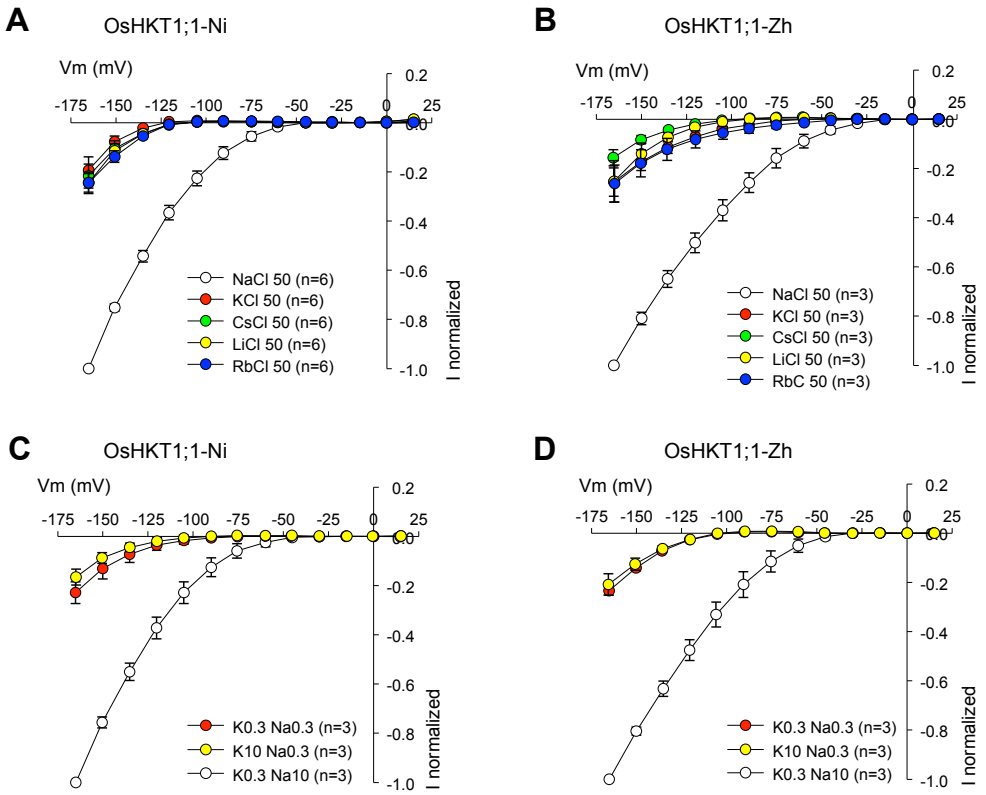


Fig S14. HKT1;1-Ni and HKT1;1-Zh display high selectivity for Na^+ against other monovalent cations. (A-D) Oocytes expressing HKT1;1-Ni (A, C) or HKT1;1-Zh (B, D) were successively bathed with solutions containing 50 mmoles•l^{-1} of different monovalent cations (Na^+, K^+, Li^+, Rb^+, or Cs^+, as chloride salts)(A, B) or different combinations of Na^+ and K^+ concentrations (as glutamate salts)(C, D). Shown I-V relationships were drawn using normalized currents (to those recorded in each oocyte at -165 mV in 50 mM Na^+ in (A) and (B), and in 10

mM Na^+ and 0.3 mM K^+ in (C) and (D)), in order to suppress small differences in expression level between oocytes. Data are means ± SE.

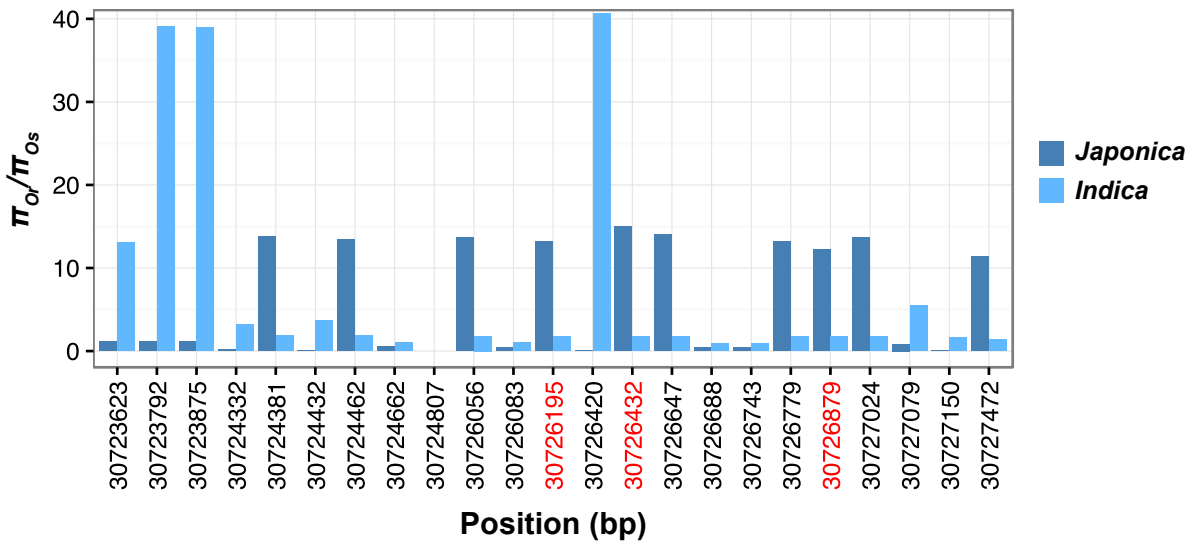


Figure S15. Comparisons of nucleotide diversity between cultivated and wild rice in the region surrounding *HKT1;1*. The three non-synonymous SNPs are highlighted in red. The y-axis represents the ratio of the nucleotide diversity (pi) in *Oryza rufipogon* to *Oryza sativa japonica* or *Oryza sativa indica*

Table S1. Mean heritability estimates for the ten traits recorded.

		Covariance Parameter Estimate						
Trait	p-val Genotype* Period	Genotype	Genotype* Period	Residual	Entry Mean Heritability	R-square	CV (%)	No. of Experimental Replicates
Root Biomass Stress	0.073	0.0083	0.0006	0.0129	0.7729	0.78	31.14	6
Root Biomass Control	0.313	0.0265	0.0005	0.0580	0.7277	0.66	34.40	6
Shoot Biomass Stress	0.017	0.0715	0.0057	0.0821	0.8123	0.81	17.91	6
Shoot Biomass Control	0.382	0.1306	0	0.1511	0.8383	0.75	17.58	6
Shoot Na Content*	-	0.0017	-	0.0031	0.5194	0.56	33.47	2
Shoot K Content*	-	0.0032	-	0.0048	0.5761	0.63	7.71	2
Shoot Na:K*	-	2395.3	-	4525.4	0.5142	0.57	35.70	2
Root Na Content	-	0.0024	-	0.0040	0.6424	0.65	17.94	3
Root K Content	-	0.0027	-	0.0167	0.3241	0.48	18.31	3
Root Na:K	-	5083.8	-	3624.5	0.808	0.76	11.94	3

Table S2: Genetic correlation analysis of root and shoot ion traits. Population structure was not accounted for in the analysis.

Trait	Root Na^+	Root K^+	Root Na^+:K^+	Shoot Na^+	Shoot K^+	Shoot Na^+:K^+
Shoot Na^+:K^+	-0.07	-0.20	0.11	0.96	-0.41	1.00
Shoot K^+	0.11	0.09	0.00	-0.17	1.00	
Shoot Na^+	-0.06	-0.21	0.12	1.00		
Root Na^+:K^+	0.81	-0.36	1.00			
Root K^+	0.25	1.00				
Root Na^+	1.00					

Table S3: Correlation of eight phenotypic traits across all subpopulations in RDP1 (*n*=383).

SB and RB indicate the ratio of biomass in salt to control for shoots and roots respectively. Highlighted cells indicate a statistically significant relationship ($p < 0.05$).

Trait	$Na^+:K^+$ Shoot	Na^+ Shoot	K^+ Shoot	$Na^+:K^+$ Root	Na^+ Root	K^+ Root	SB	RB
RB	-0.15	-0.16	0.01	0.01	-0.04	-0.08	0.55	
SB	-0.04	-0.07	-0.08	-0.08	-0.03	0.06		
K^+ Root	-0.12	-0.09	0.14	-0.32	0.47			
Na^+ Root	-0.02	0	0.14	0.66				
$Na^+:K^+$ Root	0.06	0.06	0.04					
K^+ Shoot	-0.35	-0.13						
Na^+ Shoot	0.96							
$Na^+:K^+$ Shoot								

Table S4: Correlation of eight phenotypic traits in the *admix* subpopulation (*n*=53). SB and RB indicate the ratio of biomass in salt to control for shoots and roots respectively. Highlighted cells indicate a statistically significant relationship ($p < 0.05$).

Trait	$Na^+:K^+$ Shoot	Na^+ Shoot	K^+ Shoot	$Na^+:K^+$ Root	Na^+ Root	K^+ Root	SB	RB
RB	0.13	0.16	-0.16	-0.38	-0.27	0.12	0.61	
SB	0.09	0.10	-0.15	-0.40	-0.24	0.26		
K^+ Root	0.05	0.11	0.16	-0.35	0.32			
Na^+ Root	0.01	0.01	0.33	0.73				
$Na^+:K^+$ Root	0.04	-0.02	0.16					
K^+ Shoot	-0.29	0.04						
Na^+ Shoot	0.91							
$Na^+:K^+$ Shoot								

Table S5: Correlation of eight phenotypic traits in the *aus* subpopulation (*n*=52). SB and RB indicate the ratio of biomass in salt to control for shoots and roots respectively. Highlighted cells indicate a statistically significant relationship ($p < 0.05$).

Trait	$Na^+:K^+$ Shoot	Na^+ Shoot	K^+ Shoot	$Na^+:K^+$ Root	Na^+ Root	K^+ Root	SB	RB
RB	-0.13	-0.14	0.02	-0.06	-0.14	-0.12	0.58	
SB	-0.17	-0.21	-0.05	-0.18	-0.07	0.05		
K^+ Root	-0.13	-0.11	0.09	-0.42	0.64			
Na^+ Root	0.12	0.18	0.13	0.42				
$Na^+:K^+$ Root	0.26	0.31	0.05					
K^+ Shoot	-0.28	0.01						
Na^+ Shoot	0.95							
$Na^+:K^+$ Shoot								

Table S6: Correlation of eight phenotypic traits in the *indica* subpopulation (*n*=77). SB and RB indicate the ratio of biomass in salt to control for shoots and roots respectively. Highlighted cells indicate a statistically significant relationship ($p < 0.05$).

Trait	$Na^+:K^+$ Shoot	Na^+ Shoot	K^+ Shoot	$Na^+:K^+$ Root	Na^+ Root	K^+ Root	SB	RB
RB	-0.32	-0.28	0.32	-0.12	-0.29	-0.25	0.45	
SB	-0.07	-0.09	-0.03	0.05	-0.12	-0.21		
K^+ Root	0.11	0.13	0.13	-0.19	0.58			
Na^+ Root	0.12	0.16	0.19	0.67				
$Na^+:K^+$ Root	0	0.05	0.12					
K^+ Shoot	-0.42	-0.23						
Na^+ Shoot	0.97							
$Na^+:K^+$ Shoot								

Table S7: Correlation of eight phenotypic traits in the *temperate japonica* subpopulation (*n*=91). SB and RB indicate the ratio of biomass in salt to control for shoots and roots respectively. Highlighted cells indicate a statistically significant relationship ($p < 0.05$).

Trait	$Na^+:K^+$ Shoot	Na^+ Shoot	K+ Shoot	$Na^+:K^+$ Root	Na^+ Root	K+ Root	SB	RB
RB	-0.12	-0.18	-0.17	-0.04	-0.13	-0.12	0.55	
SB	-0.12	-0.11	0.08	-0.19	-0.08	0.09		
K+ Root	-0.13	-0.04	0.27	-0.36	0.62			
Na^+ Root	0.17	0.22	0.14	0.43				
$Na^+:K^+$ Root	0.32	0.29	-0.13					
K+ Shoot	-0.14	0.07						
Na^+ Shoot	0.97							
$Na^+:K^+$ Shoot								

Table S8: Correlation of eight phenotypic traits in the *tropical japonica* subpopulation (*n*=84). SB and RB indicate the ratio of biomass in salt to control for shoots and roots respectively. Highlighted cells indicate a statistically significant relationship ($p < 0.05$).

Trait	$Na^+:K^+$ Shoot	Na^+ Shoot	K+ Shoot	$Na^+:K^+$ Root	Na^+ Root	K+ Root	SB	RB
RB	-0.26	-0.25	0.12	-0.04	-0.11	-0.03	0.46	
SB	-0.11	-0.15	-0.08	-0.24	-0.09	0.21		
K^+ Root	-0.12	-0.13	0.02	-0.52	0.42			
Na^+ Root	0.10	0.14	0.12	0.53				
$Na^+:K^+$ Root	0.16	0.22	0.11					
K^+ Shoot	-0.44	-0.20						
Na^+ Shoot	0.96							
$Na^+:K^+$ Shoot								

Table S9: List of primer sequences used for gene expression analysis and SNP confirmation. BP adaptor sequences are in red.

Gene	Direction	Primer Sequence	Purpose
LOC_Os01g51380	F	TGTTATGAATTCTATTGCCCGTGG	SNP Confirmation
LOC_Os01g51380	R	TGCTGTGGTGAGAGCAATTCT	SNP Confirmation
LOC_Os01g51420	F	ATGGACTCCTCCCGCTCCT	SNP Confirmation
LOC_Os01g51420	R	GAGCAGCTGATCTTCTTGTAGAG	SNP Confirmation
LOC_Os04g51820	F	ATGCATCCACCAAGTTTAG	SNP Confirmation
LOC_Os04g51820	R	CATTACTTGCAGTAGAATCAG	SNP Confirmation
LOC_Os04g51830	F	TGAACGCGCTGTTCCTCGC	Real-time qPCR
LOC_Os04g51830	R	CCATATGCACTGACAACTTCGA	Real-time qPCR
LOC_Os04g51820	F	TAGCCTACATTGCCGAAAGC	Real-time qPCR
LOC_Os04g51820	R	ACAAGACAAGCCGGTGAATC	Real-time qPCR
LOC_Os04g02820	F	ATTTGGTGCTTTGGACCTGA	Real-time qPCR (Internal Control)
LOC_Os04g02820	R	TGAATAGCATCAGCGTGGAG	Real-time qPCR (Internal Control)
LOC_Os04g51830	F	TGAACGCGCTGTTCCTCGC	Real-time qPCR
LOC_Os04g51830	R	CCATATGCACTGACAACTTCGA	Real-time qPCR
LOC_Os04g51820	F	AAAAAGCAGGCTGAGTATGCTTACATCCTTC	HKT1;1 RNAi
LOC_Os04g51820	R	AGAAAGCTGGGTCATTACTTGCAGTAGAATCAG	HKT1;1 RNAi
LOC_Os04g51830	F	caccACCGTCCTGATGCTCC	HKT1;4 RNAi
LOC_Os04g51830	R	GATCTGTCCATGGCGTCGC	HKT1;4 RNAi
LOC_Os04g51820	F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGTTTAAAACC GCTGTGGAA	HKT1;1 native overexpression

Supplemental File S2. QTL locations for the eight phenotypic traits and the percent variation explained by the most significant SNP for each QTL.

Supplemental File S3. Transcripts within *RNC4* displaying differences in expression between allelic groups at SNP-4-30535352

Supplemental File S4. High-confidence genetic variants (at least 10x coverage) found in the ORFs of genes within *RNC4*.

CHAPTER 3

IMAGE HARVEST: AN OPEN SOURCE PLATFORM FOR HIGH-THROUGHPUT PLANT IMAGE PROCESSING AND ANALYSIS

Avi C. Knecht, Malachy T. Campbell, Adam Caprez, David R. Swanson, Harkamal Walia

Published: May 3 2016; https://doi.org/10.1093/jxb/erw176

Abstract

High-throughput plant phenotyping is an effective approach to bridge the genotype-to-phenotype gap in crops. Phenomics experiments typically result in large-scale image datasets, which are not amenable for processing on desktop computers, thus creating a bottleneck in the image analysis pipeline. Here, we present an open-source, flexible image-analysis framework, called Image Harvest (IH), for processing images originating from high-throughput plant phenotyping platforms. Image Harvest is developed to perform parallel processing on computing grids and provides an integrated feature for metadata extraction from large-scale file organization. Moreover, the integration of IH with the Open Science Grid provides academic researchers with the computational resources required for processing large image datasets at no cost. Image Harvest also offers functionalities to extract digital traits from images to interpret plant architecture-related characteristics. To demonstrate the applications of these digital traits, a rice (*Oryza sativa*) diversity panel was phenotyped and genome-wide association mapping was performed using digital traits that are used to describe different plant ideotypes. Three major quantitative trait loci were identified on rice chromosomes 4 and 6, which co-localize with quantitative trait loci known to regulate agronomically important traits in rice. Image Harvest is

an open-source software for high-throughput image processing that requires a minimal learning curve for plant biologists to analyze phenomics datasets.

3.1 Introduction

With the advent of next-generation sequencing technologies and development of high-density genotyping platforms for many crop species, the volume of genotypic data being generated has increased exponentially. However, the collection of corresponding phenotypic data has lagged and remains a laborious task, often involving destructive measurements. Large-scale phenotypic (phenomics) data for traits of interest is essential to realize the full potential of the enormous progress in crop genomics. The generation of phenomics data for plants is challenging because many of the commercially valuable traits are quantitative, exhibit high variability across diverse environments, and change throughout the plant's life cycle, thus adding to cost and labor. Further, many phenotypic traits are not amenable for visual scoring and observations on a large-scale. These challenges have led to a translational bottleneck, often referred to as the genotype-phenotype gap. This challenge is discussed in several recent reviews (Furbank and Tester, 2011; Sozzani and Benfey, 2011; Gjuvsland et al., 2013; Yang et al., 2013; Brown et al., 2014; Fraas and Luthen, 2015).

To bridge this gap, several efforts in the public sector and industry have started to materialize through the establishment of research centers focusing on the development and applications of high-throughput phenomics technologies (Yang et al., 2013). Image-based plant phenomics is one area becoming increasingly accessible in the public domain (Australian Plant Phenomics Facility, European Plant Phenotyping Network, PHENOME: The French Plant Phenomics Network (FPPN), High-Throughput Rice Phenotyping Facility (HRPF; Huazhong University, China), Donald Danforth Plant Science Center, and University of Nebraska-Lincoln Innovation

Campus, LeasyScan at ICRISAT). These high-throughput imaging platforms allow for non-destructive measurements to be recorded accurately and frequently during the course of an experiment. Such systems utilize a series of cameras to estimate plant growth, temperature, water content or chlorophyll characteristics on a large-scale, and provide researchers with valuable insight into dynamic physiological changes occurring during the course of plant development or in response to environmental changes. These platforms are often fully automated and therefore have lower technical errors compared to traditional phenotyping methods. Overall, the optics-based systems greatly enhance the ability to capture quantitative traits on a temporal and spatial scale, and through the integration of mechanical automation, make high-throughput plant phenotyping accessible to public researchers.

The generation of large-volume image data currently is, and will, remain a constant for most plant phenomics experiments. There are two main considerations for analysis of image-based data. The first is access to image analysis tools to extract biologically meaningful digital traits. The second consideration is image storage, standardized cataloging protocols, and accessibility through public repositories so it is widely available for the plant research community. Publically accessible tools for image analysis and image storage (Bisque/iPlant) have emerged in recent years, though this field still remains in infancy (Kvilekval et al., 2009; Yazdanbakhsh and Fisahn, 2009; Galkovskyi et al., 2012; Tanabata et al., 2012; Pound et al., 2013; Brown et al., 2014; Klukas et al., 2014; Whan et al., 2014; Fahlgren et al., 2015; Müller-Linow et al., 2015). However, this challenge is largely tractable because many of the image analysis resources that are publically available in the field of computer vision can be adapted for plant phenomics data. One such resource is OpenCV, which is a library of programming functions for processing many types of images (Bradski, 2000). However, adapting and utilizing these functionalities for plant

phenomics requires computational and programming expertise, thus limiting wider utilization by the plant science community. This hurdle has been addressed with several timely resources, such as PlantCV, IAP, and others (Yazdanbakhsh and Fisahn, 2009; Galkovskyi et al., 2012; Tanabata et al., 2012; Pound et al., 2013; Brown et al., 2014; Klukas et al., 2014; Whan et al., 2014; Fahlgren et al., 2015; Müller-Linow et al., 2015).

Klukas et al. and Fahlgren et al. have both developed software for extracting biological information from images captured with LemnaTec platforms (http://www.lemnatec.com) (Klukas et al., 2014; Fahlgren et al., 2015). Despite the power and flexibility provided by these programs, the loading and collection of experimental metadata from large datasets is done manually or through individual scripts developed by various research groups, making these tasks time consuming and laborious. To put this in context, the number of files generated for a triplicated medium-scale experiment with 3 weeks of multi-camera imaging and ~300 genotypes can easily exceed three million. This can make even simple file management tasks computationally challenging for plant biologists. Moreover, the execution of image processing tasks is computationally intensive and requires access to high-throughput computational grids, and a proficiency in one or more programming languages. Many of the available plant image analysis software cannot easily be adapted by plant biologists for implementing processing workflows on computational grids, and require extensive modifications to fully realize the potential of distributing intensive computational tasks across an array of processors.

To address some of these challenges, we have developed Image Harvest (IH) as an open source and flexible framework for making high-throughput image analysis accessible for the plant biologists. Image Harvest (IH) is an open-source Python library for processing and analyzing plant images. In addition to providing powerful functions for processing several types of images

(conventional color and fluorescence images), IH contains functions that greatly simplify the collection of metadata from the organizational structure of raw image databases. IH provides an option for implementing basic statistical functions and provides several definitions of digital traits to describe plant growth, morphology, and physiological responses. Moreover, IH has been integrated with the Open Science Grid, which provides grid computing resources to academic users at no cost, thus reducing the overhead costs associated with the running and analyzing phenomics experiments. Here, we detail the image analysis functionalities of Image Harvest and demonstrate the value of some of the plant architecture-related digital traits by mapping them to the rice genome. The specific aims of this manuscript are to (1) describe the creation and execution of processing workflows on a local machine and on the high-throughput computational cluster, (2) demonstrate the processing power and accuracy of IH, (3) and present some of the downstream applications of image-derived digital traits from two crop species.

3.2 Results

3.2.1 Software Overview

Image Harvest (IH) is an open-source Python library for processing and analyzing plant images. IH provides integrated functions for complete image analysis from processing to descriptive statistics. The software is written in Python and uses algorithms from OpenCV to extract plant objects from complex images. SciPy is used for basic downstream statistical analysis of digital traits (Jones et al., 2007).

The implementation and execution of image processing steps varies depending on the nature and size of the input files. IH is available as standalone software for OS X, Windows and Linux, and is also compatible for high-throughput image processing on computing clusters. The standalone

software allows users to develop processing pipelines and process a small number of images. Once processing pipelines have been defined, they can be scaled up to process thousands of images in a high-throughput computing environment. IH workflows have been integrated into the Open Science Grid, (http://www.opensciencegrid.org), which is comprised of ~100,000 opportunistically available processors (Pordes et al., 2007). The presence of these workflows on OSG provides high throughput computing resources necessary for processing large image sets in a matter of hours for the plant science community. Although several software tools are available to perform image analysis, the ability to create distributed computing workflows is one of the salient and unique features of IH and is currently not implemented in other software.

3.2.2 Image Capture

IH is a flexible platform, capable of processing images from various sources, ranging from automated imaging platforms to conventional hand-held cameras. The input images require the retention of straightforward parameters to ensure optimal processing of images and accuracy of digital traits. Imaging should be done in a uniform/controlled environment. It is recommended that plants be imaged in an environment with multiple light sources to minimize shadows and provide a uniform background for images (Supplemental Fig S1). The plant should be placed in front of a uniform background that is very different in color than that of the plant (for instance white, blue or black). Based on our preliminary experiments, blue or white backgrounds are easiest to differentiate background features from plant pixels. These colors are also suitable for pots or frames that support the plant. If comparing multiple images, the camera and plant position should be fixed, and lighting conditions should be consistent from image to image. Slight movement caused by mechanical agitation or airflow may cause blurring, which reduces accuracy, and therefore the user should ensure that movement is minimized. Lastly, the imaging

environment should be clear of plant matter, soil or other debris, as these have similar colors properties to the plant and may be classified as “plant pixels” during processing.

3.2.3 Developing a processing workflow on a local computer

Prior to implementing a computationally intensive workflow on thousands of images, an appropriate workflow should be developed on a local computer using a few representative plant images. IH allows for the results of each processing step to be monitored in real-time. Here we provide a simple processing pipeline for extracting plant pixels from a single color (RGB) image of a rice plant (*cv* 9311) at maturity (Fig. 1A). The images were captured using a conventional SLR camera (Canon Rebel T1i) and were imaged in a homemade imaging chamber to provide adequate lighting (Supplemental Fig. S1). The raw images, as well as the full processing script is provided at the IH website (http://cropstressgenomics.org/data/html/ex_script_camera2.html).

Although workflows are specialized, there are common steps and design decisions in each workflow with wider applicability. For all image types, processing can be separated into four distinct modules: (1) background removal, (2) noise reduction, (3) cropping, and (4) data gathering. A list of functions available for each of the four classes of processing tasks is provided as Supplemental Table S1. The purpose of background removal is to remove a majority of the pixels that can clearly be defined as background, whereas noise reduction is fine-tuned for specific areas that may contain both background and plant pixels, such as the plant container. Finally, cropping is used to extract only the plant pixels from the image and, if performed effectively, increases data reliability.

3.2.3.1 Background Removal

The first image processing step is a simple cropping function (`crop()`) to remove features outside of the background at the edges of the image. This step provides a more uniform background and allows the plant object to be extracted from the background using a few simple processing functions. In practice, the camera and plant positions will be fixed, so once parameters have been optimized for this function from a single image, the same can be applied to all other images. Next, the `colorFilter()` function is used to retain pixels with color qualities that satisfy a logical argument. Based on our logic, IH retains the pixels that meet the following condition: the green value of the pixel must exceed the blue channel value. `colorFilter()` can also be applied to a specific area of image by providing an optional region of interest (ROI) in the form [ystart, yend, xstart, xend]. These simple functions are sufficient to remove the majority of the background from the image.

3.2.3.2 Noise Reduction and image cropping

The `crop` and `colorFilter` functions remove the majority of background pixels, however, other small background features remain in the image and must be removed through additional processing steps. These additional features are removed from the images based on their size. The `contourChop()` function is used to remove these small background features. The `contourChop()` function takes two arguments: the image to be processed and a minimum threshold area for contours to be removed.

Alternatively, IH has several morphological operations that can be used in place of `contourChop()` to remove small non-plant objects from the image. For example, morphological opening erodes all borders of the objects in the image and subsequently adds

borders back to the shapes remaining after the erosion step. Thus, much of the noise in the background of the image can be removed.

The final step of image processing is to crop the image to just the plant to obtain an accurate measure of plant height and width. To do this, IH first generates a binary image and uses the `contourCut()` function to crop the image as a final step. The image is cropped so all contours that are greater than the specified area are included in the final output image.

3.2.4 Executing processing workflows on a computing cluster

Once the user has established an appropriate processing workflow, the processing pipeline can be scaled to large, high-throughput data sets (> 50,000 images) and executed on a high-throughput computing grid (Fig 2). For datasets of this size, the execution of basic processing tasks is challenging. IH utilizes Pegasus for translating a series of computational tasks into a Directed Acyclic Graph and uses HTCondor to execute the jobs in parallel on a computing cluster (Thain et al., 2005). Prior to initiating any workflow, the user should ensure that the proper environmental variables are established and Pegasus and HTCondor are installed on the computing grid. Alternatively, users may consider the OSG Connect service, which has IH and all the prerequisites installed, and is free to all academic users (Pordes et al., 2007). Full-scale datasets are available at /stash/project/@RicePhenomics/public/. It is recommend that a structured directory be created that provides a detailed classification for each image (i.e. species, treatment, imaging date). The structure and organization of images are essential for the proper loading and creation of metadata.

The successful execution of workflows for processing large datasets relies on the creation of three JSON (JavaScript Object Notation) formatted template files: Image Loading, Image Processing, and Configuration. Below, a description and purpose for each template file is

provided. The aim of this section is not to describe each processing step, as this has been discussed in the previous section, but to provide a brief overview of the necessary steps to execute workflow processing on computing clusters. An example of the necessary template files can be accessed through the IH website (http://cropstressgenomics.org/data/html/ex_workflow_1.html).

3.2.4.1 Image Loading

For large image sets the loading and the creation of descriptors for each image can be a complex and laborious task. To circumvent these issues, IH uses an automated crawling process, which parses the information captured in the hierarchical structure and names of the directories to load metadata for images. The image loading step in the IH workflow takes care of two key problems posed by having large data sets, assuming the template file is written correctly. First, it tracks all images, and loads them into a single database, making it easy to locate and process files in the future. Second, all meta-data definitions are done as a result of the crawling.

The image loading template file will vary significantly based on how the images are named and organized. However, IH's crawling function is designed to be flexible enough to be adapted to most directory structures. In this example, the images were captured with a LemnaTec Scanalyzer 3D and the pot identification number as well as the image time stamp are included in the directory names. Within each time stamped folder there are several subdirectories that are named according to the image perspective (top view or side view) and image type (VIS/RBG or FLUO). It is assumed, if using a different system, that an appropriate method is in place to efficiently store and organize images after capture, as should be the case with other phenomics platforms that are capable of capturing thousands of images. Basic information such as the plant identification number, timestamp information and image type can be extracted from the directory

naming and structure. Additional metadata, such as genotype names, replicate number, position in the greenhouse, etc. can be appended to the final results file by supplying a comma-separated file. The file should contain a key column, which matches the identifiers extracted from the directory names. Several examples of IH's automated crawling are provided at the IH webpage (http://cropstressgenomics.org/data/). This allows the final results file to be navigated and mined with greater ease and simplifies downstream statistical analysis.

3.2.4.2 Processing

The image-processing template designates the type and order of image processing functions. Separate workflows should be designed for each image type (rgbsv, rgbtv, fluosv, etc.) and, as a result, the template may become quite long. Each workflow is defined as a list of jobs. The jobs are not necessarily executed in the order specified in the list, but the job definition should be structured such that it follows the list order as closely as possible for ease of reading. It is essential for the name of each workflow to match the name of the image types in the database.

3.2.4.3 Execution of IH and Statistics Workflows

The execution of complex workflows in IH is relatively easy provided the proper template files (Image Loading, Image Processing, and Configuration) and the directories have been established, and are fulfilled using the `ih-run` command. This creates a date and time stamped directory, which contains the output database, a copy of the submitted templates and all non-image raw-input files. IH provides several functions for extracting digital traits that describe plant morphology, growth and color/spectral qualities from processed images. A complete list of available functions is provided as Supplemental Table S1. These metrics can be combined from multiple views to describe different morphological traits.

To demonstrate the performance of IH we executed a processing workflow for FLUO side view and VIS side and top view images using several image sets of varying size (77665, 38357 and 19457 images). Images were processed on a computing cluster (106 node LINUX cluster, Opteron 6272 2.1GHz, 4 CPU per node, 256 GB RAM per node). The small and medium image sets were completely processed in less than 13 hours, while the large image set consisting of > 77K images were processed in slightly more than 30 hrs (Fig. 3). The total time (from the user's end) that is require to process image sets on a computing cluster is dependent on the number of tasks and the number of available processors. For the small and medium datasets (19457 and 38357 images, respectively), the number of available processors was likely greater than the number of tasks, which explains the nearly equal processing times. However, in the large dataset the number of available processors was far fewer than the number of tasks.

For all image sets the actual image processing tasks accounted for less than 1% of the total processing times. Tasks involved with trait extraction and data compilation accounted for the majority of computation time. The function `"ih-extract-multi"` accounted for between 57-69% of the total computation time, while `"ih-sql-aggregate"` function accounted for between 25-31% of total computation time. The `"ih-extract-multi"` is used to extract multiple metrics (color data, moment data, dimensions etc.) that are chosen by the user from processed images. A full list of available metrics is provided as Supplemental Table S2. While the computation time for simple morphological traits is typically low, the color classification/histogram function implemented in this workflow requires considerably more time and is likely the major factor contributing to the large computation times for this task. IH uses Sqlite to manage processing outputs, which improves portability but requires a relatively time-intensive collation step. Although IH can extract digital metrics from multiple images, without

running into concurrency issues, the data must be written to different databases at the worker nodes, which then must be aggregated into a single database. The `"ih-sql-aggregate"` function combines output from all individual databases to a single final database.

3.2.5 Assessing plant growth with IH

IH contains various functions for extracting digital traits describing plant shape, size and color/spectral properties from the final plant object. A subset of these descriptors is illustrated in Fig. 1B, and the full list is available as Supplemental Table S1. Several software have been developed to process plant image data (Yazdanbakhsh and Fisahn, 2009; Galkovskyi et al., 2012; Tanabata et al., 2012; Pound et al., 2013; Brown et al., 2014; Klukas et al., 2014; Whan et al., 2014; Fahlgren et al., 2015; Müller-Linow et al., 2015). To compare the accuracy of IH with existing plant image analysis software, a dataset consisting of 144 RGB images of from 72 rice plants during the early vegetative growth phase were processed using IH, LemnaGrid (www.LemnaTec.com) and PlantCV (Fahlgren *et al.*, 2015). Plants were imaged from two side view angles and the foreground/plant pixels were summed from both side angles, and were used as a proxy for shoot biomass. This digital representation of shoot biomass, here termed projected shoot area (PSA), was compared to three manual phenotypic measurements (shoot area, shoot fresh weight and shoot dry weight) recorded from the same group of plants. All software exhibited very high correlation with all phenotypic measurements (Fig. 4). Overall, the differences in accuracy among the three processing software products were minor. LemnaGrid displayed slightly higher accuracy than both IH and PlantCV. Both LemnaGrid and PlantCV displayed slightly higher correlation with shoot area (r^2=0.94) compared to IH (r^2=0.93). The results indicate that the accuracy of IH is comparable to other image processing software.

3.2.6 From pixel-based digital traits to genes: A case study in rice

Selection for morphological traits, such as plant height, lodging resistance and tillering capacity, are major contributing factors for increasing rice productivity over the past century. Rice breeders have defined specific ideotypes, which combine several traits to maximize photosynthetic capacity, growth habit and grain production (Tandon and Jain, 2004; Peng et al., 2008). However, classification and selection of ideotypes are highly subjective, and based on visual classification systems developed by plant breeders. The advent of image-based plant phenomics allows for the nondestructive evaluation of phenotypes while reducing human error, and provides an opportunity to define digital traits that describe multidimensional phenotypes and morphological features. In this section we sought to identify the major plant architectural classes from rice images, define digital traits that describe aspects of plant morphology and demonstrate the potential to identify genomic regions associated with these digital traits. To this end, we imaged a panel of 376 diverse rice genotypes during the late tillering stage (41 days after transplant) using a RGB/VIS camera.

To identify the major shape/plant architectural types in rice germplasm, raw image moments (M) were extracted from 1,548 side view (SV) and 774 top view (TV) images and hierarchical clustering analysis was performed using the adjusted mean values for these digital metrics. Raw image moments extracted from binary images provide a numerical description of various properties of the foreground object, such as center of mass or orientation. Hierarchical clustering of raw image moments identified three major phenotypic groups (Fig. 5A). To determine how these groups differed in morphological features, a one-way ANOVA was performed using 23 digital traits that describe various morphological qualities (Fig. 5B; Table S3; Table S4). Significant differences between the groups ($p < 0.0001$) were observed for 20 traits that describe

plant width, compactness and biomass (Fig. 5B; Table S4). Interestingly, no significant differences were observed for traits corresponding to plant height, suggesting that these groups largely differ in biomass production and growth habit.

To identify genomic regions associated with plant architecture, genome wide association mapping was performed using 36,901 SNPs (single nucleotide polymorphisms) and the 23 digital traits for 359 genotypes (Supplemental Fig. S2; Fig S3). Three highly significant peaks were identified for $Den1_{TV}$ and GH1, which may be proxies for canopy density and erect growth habit respectively (Fig. 6). Notably, the most significant SNPs co-localize with genes previously associated with plant growth and development. A rice gene *OsRCN4* (LOC_Os04g33570) similar to *TERMINAL FLOWER1* (*TFL1*) in Arabidopsis was located approximately 25kb upstream of the most significant GWAS peak on chromosome 4 (Supplemental File S1). Moreover, several QTLs/genes known to regulate tillering, flowering time and hormone homeostasis (Table 1) have been mapped to the long arm of rice chromosome 6, where a number of highly significant SNPs populated a region of ~861 Kb. The identification of significant associations with regions known to regulate plant morphology and flowering time suggest that digital trait outputs from IH can be a useful tool for the evaluation and selection of genotypes exhibiting optimal morphological and agronomic traits.

3.3 Discussion

Plant phenomics is a new and powerful approach, however there are still considerable challenges for plant scientists to extract biologically meaningful information from images. Many existing algorithms have been developed in the field of computer vision and can be utilized to extract morphological and color properties from plant images. However, an extensive knowledge of several programming languages is necessary for the efficient implementation of these methods

on large data sets, and may be a skill set that is rare in most plant biology research groups. To allow greater accessibility to these powerful algorithms in computer vision, IH combines utilities from OpenCV and SciPy, and provides a user-friendly, broadly applicable image-processing framework that can be run locally on conventional desktops to develop processing pipelines. Processing functions available in IH can be applied to images captured from several sources, such as handheld SLR cameras or phenomics platforms other than LemnaTec, and can be used to process various types of images (RGB and FLUO).

In recent years several groups have released software to process images derived from phenomics platforms (Klukas et al., 2014; Fahlgren et al., 2015). IH has several advantages over existing software. First, IH is an open source software. While other proprietary software, such as LemnaGrid, work seamlessly with LemnaTec platforms, the user is required to purchase software licenses and is often confined to the processing algorithms and metrics offered within the software package. Although LemnaGrid offers numerous descriptive metrics, users often want to address biological processes beyond simple morphology, which may cause users to seek resources outside of LemnaGrid. For example, a binning strategy was used by Campbell *et al.* to identify color ranges in fluorescence images that were responsive to salinity treatment (Campbell et al., 2015). At the time of publication, these methods implemented were not available in LemnaGrid. Other open-source software offers numerous metrics to describe plant morphology and color properties, and also proved a means for new algorithms to be added, thus improving the flexibility of the software to address specific research needs (Klukas et al., 2014; Fahlgren et al., 2015).

Second, IH can process images generated from various sources, included other non-LemnaTec phenomics platforms, flatbed scanners, SLR cameras, and non-RGB images (fluorescence,

hyperspectral, etc.). In addition to the pipelines presented above, several tutorials are provided at the IH website (http://cropstressgenomics.org/data/), which provide a complete description of pipelines for processing FLUO and RGB images, phenomics platforms that use overhead cameras to phenotype trays of plants, and quantifying morphological traits from images of seeds obtained with a conventional desktop scanner. Once a pipeline has been developed, the modular organization of IH functions allow processing tasks to be easily translated to supercomputer workflows. In contrast, at the time of this publication, the functionalities of IAP are not easily accessible for images generated from platforms other than LemnaTec.

Third, IH has been developed to utilize the power of grid computing. Phenomics experiments produce a large amount of data, and although several user-friendly processing software options are available, the computational requirements for such dataset exceed the capacity of desktop computers or workstations. Access to linux-based computational clusters and a working knowledge of computer programing to integrate several existing software packages is necessary to process large image sets. IH circumvents these requirements by combining functionalities from OpenCV, SciPy, Condor, Sqlite and Pegasus into a single toolkit that can be accessed by plant biologists with minimal programming expertise. Moreover, IH has been developed to be highly portable; allowing simple installation to most Linux based computing grids with minimal configuration and assistance from system administrators.

The integration of IH with OSG provides the computational resources necessary for processing large image sets in a reasonable time. At the time of this publication, several software packages for processing images derived from phenomics platforms are publically available that have capabilities for parallelizing workflows. For example PlantCV uses a terminal multiplexer, such as Tmux (https://tmux.github.io/) or Screen, to submit multiple processing jobs in parallel.

However, these methods used for parallelization may not be adequate for realizing the full potential of super computing facilities. Using this type of approach, computation is limited to the resources of a single machine. By contrast, with IH individual processing tasks within a workflow can be distributed to dozens or even hundreds of machines spread across the country in a manner that is transparent to the user, thus facilitating quick parallelization of processing tasks and analyses. This effectively allows any academic research group to develop and implement a processing workflow for large datasets without the need to purchase dedicated servers or workstations.

Image-based phenomics allows physiological and morphological traits to be quantified nondestructively at a high frequency throughout a given period. The images contain important digital traits that may provide phenotypes beyond what is typically quantified in conventional manual phenotyping. For instance, in the field of computer vision image moments are used to describe various properties of a contour, such as the size or centroid of the object. Image moments have been used in plant and mammalian systems to classify various or describe features from images (Ortiz et al., 2013; Rojo-Arreola et al., 2014; Lin et al., 2015) . The applicability of IH functions for phenotyping large mapping populations are highlighted in the GWAS presented above where 23 digital growth-related traits were quantified for 359 genotypes. The most highly significant signals were for metrics used to describe canopy density from top view images ($Den1_{TV}$, $Den3_{TV}$) and GH1, which are traits that would be difficult to quantify manually for a population this size. Moreover, these SNPs were in close proximity to genes that are known to regulate plant morphology in rice or other species, indicating that these signals are biologically relevant. For instance, an *AtTFL1* ortholog was identified approximately 25 kB upstream of the most significant SNP associated with GH1, GH2, GH3, and GH4 on chromosome 4. While the

role for this gene in rice is not known, *AtTFL1* represses the transition from vegetative to reproductive development in Arabidopsis (Shannon and Meeks-Wagner, 1991; Wickland and Hanzawa, 2015). In rice, ectopic overexpression of the *AtTFL1* orthologs *RCN1* and *RCN2*, prolong the vegetative growth phase and lead to an over-production of leaves (Nakagawa et al., 2002). Although *OsRCN4* has yet to be characterized, the digital phenotypes exhibited by the allelic groups at the QTL on chromosome 4 are in agreement with the phenotypes exhibited by other *TFL1* homologs in rice.

In addition to *TFL1*, two genes were identified within QTLs that have been shown to regulate plant growth and morphology through hormones such as jasmonic acid (JA) and auxin. A QTL spanning ~400 kb on chromosome 4 was associated with four digital traits ($Den1_{TV}$, $Den3_{TV}$, $Den2_{SV}$, PSA). *OsRJR1* (*LOC_Os04g23550*) encodes a bHLH transcription factor and is located less than 2kb upstream of the most significant SNP for $Den3_{TV}$ (id4003991). A study by Kiribuchi *et al.* showed that *OsRJR1* is a positive regulator of JA-mediated inhibition of shoot growth (Kiribuchi et al., 2004). While this study provided biological insight into *OsRJR1* function, further experimentation is required to elucidate the role of this gene in natural variation of plant morphology. A second gene known to be involved with auxin signaling was identified within a QTL associated with $Den1_{TV}$ and $Den3_{TV}$ on chromosome 12. *OsIAA3* is a member of the *Aux*/*IAA* transcription factor family, and has been shown to regulate auxin responses in rice (Nakamura et al., 2006). Overexpression of *OsIAA3* resulted in auxin insensitivity, with transgenic plants exhibiting reduced crown root formation, shorter leaf sheaths and abnormal leaf development, indicating that it is a negative regulator of auxin responses (Nakamura et al., 2006). The identification of genes that are known to regulate morphology and development in

rice and other species demonstrate that these digital traits can be may be associated with important biological processes.

Over the past decade the amount of genotypic data has increased substantially which has allowed researchers to examine the genome and transcriptome at a high resolution and associate genomic data with a phenotype. However, phenotyping remains a significant bottleneck for bridging the genotype-phenotype gap. However, the advent of image-based phenomics allows large populations of plants to be phenotype non-destructively and provides an opportunity to quantify phenotypes that were traditionally evaluated subjectively or those that are not visible with the human eye. Despite the large investment in the construction of phenomics facilities, few open source software solutions have been developed that are capable of processing large datasets in a reasonable time. IH provides a framework to develop processing workflows and execute them in a distributed computing infrastructure.

3.4 Materials and Methods

3.4.1 Plant materials

This study included 376 of the 421 original RDP1 accessions (Zhao et al., 2011; Famoso et al., 2011; Eizenga et al., 2014). Forty-five accessions were not included due to lack of seed availability or poor seed quality. Accessions were obtained from the USDA-ARS Dale Bumpers Rice Research Center and purified through single seed descent prior to phenotyping.

3.4.2 Plant Growth Conditions

Seeds from 373 genotypes from the rice diversity panel were surface sterilized with fungicide, Thiram®, and germinated on moist paper towels in plastic boxes for three days (Zhao et al., 2011; Famoso et al., 2011). Three uniformly germinated seeds of each genotype were

transplanted to pots (150mm diameter x 200 mm height) filled with 2.6 kg of UC Mix. Square containers were placed underneath the pots to provide adequate water to saturate the soil. Plants were thinned to one seedling per pot six days after transplanting (DAT). For the first 20 DAT each pot was watered daily with ~150 mL from the top of the pot. Over the course of the two experiments, the greenhouse temperatures during the day averaged 28.8°C (±2.02°C, SD) and 26.0°C (±1.01°C, SD) at night. Relative humidity was maintained at 63.4% (±9.04%, SD) during the day and 69.7% (±1.73%, SD) at night (Rotation Atomiser Defensor ABS3, Condair Ltd., Pfäffikon, Schwyz, Switzerland). Twenty-one DAT, each pot was watered to a uniform weight so that approximately 600 ml of water was maintained in the soil. The experimental design was identical to that described by Campbell *et al*. (Campbell et al., 2015). Each pot was imaged daily using the LemnaTec Scanalyzer 3D system at the Plant Accelerator facility in Adelaide Australia for 18 days using two 5 megapixel RGB/VIS cameras (Basler Pilot piA2400-12gc) from three perspectives consisting of two side view angles separated by 90° and a single top view. The entire data set of can be accessed through the iPlant Collaborative upon publication. (http://mirrors.iplantcollaborative.org/browse/iplant/home/shared/walia_ rice_salt)

3.4.3 Phenotypic data analysis

Data were combined across experiments, and a linear model was fitted to calculate the adjusted means for each individual accession using the 'lsmeans' function in the LSMeans package in R (R Core Team, 2014; Lenth and Hervé, 2015). In the linear model, experiment is considered a fixed effect and accession as a random effect. The adjusted means were used for hierarchical clustering of raw moments using the complete-linkage method with Euclidian distance as the distance metric. For the comparison of 22 digital traits with 3 manual phenotypic measurements,

Pearson correlation analysis was was done using the rcorr function in the Hmisc package in R (Harrell, 2014; R Core Team, 2014).

3.4.4 Genome-wide association mapping

All accessions were genotyped using 44,000 SNPs (Zhao et al., 2011). A conventional mixed-model genome-wide association analysis was used to identify genomic regions associated with each of the 22 digital traits. The implemented mixed linear model can be summarized as: $y = X\beta + C\gamma + Zu + e$, where β and γ represent coefficient vectors for SNP effects and subpopulation principal components, respectively, which are fixed effects, u is a random effect that accounts for population structure and relatedness, Z represents the corresponding design matrices, and e is the random error term. The model was implemented using EMMA in R using a minor allele frequency cutoff of 0.05 (Kang et al., 2008; R Core Team, 2014).

3.5 Availability and requirements

Project name: Image Harvest

Project home page: http://cropstressgenomics.org/data/html/index.html; https://git.unl.edu/aknecht2/ih/

Operating system(s): Windows, Linux, OS X

Programming language: Python

Other requirements: Matplotlib 1.3.1+; NumPy 1.7+; OpenCV 2.4.x; GTK2 (including headers) 2.24+; blas 3.2.1+; lapack 3.2.1+; libav 0.5.3+; cmake 2.8.12+; ffmpeg 2.0.5+; Pandas 0.13.1+; Cython 0.21.1+; python-dateutil 1.4.1+; PyMeanShift 0.2.0+; SciPy 0.13.1+; Pegasus 4.4.0+; Java 1.8+; Perl 5.10+; HTCondor 8.3.0+

License: GNU GPL (https://git.unl.edu/aknecht2/ih/blob/master/GPL.txt)

Any restrictions to use by non-academics: *See GPL*

3.6 References

Bradski, G. 2000. The OpenCV Library. Dr. Dobb's J. Softw. Tools 25(11): 120–126.

Brown, T.B., R. Cheng, X.R.R. Sirault, T. Rungrat, K.D. Murray, M. Trtilek, R.T. Furbank, M. Badger, B.J. Pogson, and J.O. Borevitz. 2014. TraitCapture: Genomic and environment modelling of plant phenomic data. Curr. Opin. Plant Biol. 18(1): 73–79Available at http://dx.doi.org/10.1016/j.pbi.2014.02.002.

Campbell, M.T., A.C. Knecht, B. Berger, C.J. Brien, D. Wang, and H. Walia. 2015. Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. Plant Physiol. 168(August): pp.00450.2015.

Duan, M., Z. Sun, L. Shu, Y. Tan, D. Yu, X. Sun, R. Liu, Y. Li, S. Gong, and D. Yuan. 2013. Genetic analysis of an elite super-hybrid rice parent using high-density SNP markers. Rice 6(1): 1–15Available at http://dx.doi.org/10.1186/1939-8433-6-21.

Eizenga, G.C., M. Ali, R.J. Bryant, K.M. Yeater, A.M. McClung, S.R. McCouch, and others. 2014. Registration of the rice diversity panel 1 for genomewide association studies. J. Plant Regist. 8(1): 109–116.

Fahlgren, N., M. Feldman, M.A. Gehan, M.S. Wilson, C. Shyu, D.W. Bryant, S.T. Hill, C.J. McEntee, S.N. Warnasooriya, I. Kumar, T. Ficor, S. Turnipseed, K.B. Gilbert, T.P. Brutnell, J.C. Carrington, T.C. Mockler, and I. Baxter. 2015. A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in Setaria. Mol. Plant 8(10): 1520–1535Available at

http://dx.doi.org/10.1016/j.molp.2015.06.005.

Famoso, A.N., K. Zhao, R.T. Clark, C.-W. Tung, M.H. Wright, C. Bustamante, L. V Kochian, and S.R. McCouch. 2011. Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7(8): e1002221.

Fraas, S., and H. Luthen. 2015. Novel imaging-based phenotyping strategies for dissecting crosstalk in plant development. J. Exp. Bot. 66(16): 4947–4955.

Furbank, R.T., and M. Tester. 2011. Phenomics - technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 16(12): 635–644Available at http://dx.doi.org/10.1016/j.tplants.2011.09.005.

Galkovskyi, T., Y. Mileyko, A. Bucksch, B. Moore, O. Symonova, C. a Price, C.N. Topp, A.S. Iyer-Pascuzzi, P.R. Zurek, S. Fang, J. Harer, P.N. Benfey, and J.S. Weitz. 2012. GiA Roots: software for the high throughput analysis of plant root system architecture. BMC Plant Biol. 12: 116.

Gjuvsland, A.B., J.O. Vik, D.A. Beard, P.J. Hunter, and S.W. Omholt. 2013. Bridging the genotype-phenotype gap: what does it take? J. Physiol. 591(8): 2055–66Available at http://www.ncbi.nlm.nih.gov/pubmed/23401613%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3634519.

Harrell, F.E. 2014. Hmisc: Harrell Miscellaneous.

Jones, E., T. Oliphant, P. Peterson, and others. 2007. SciPy: Open source scientific tools for Python, 2001--. URL http//www. scipy. org 73: 86.

Kang, H.M., N. a Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178(3): 1709–23.

Kiribuchi, K., M. Sugimori, M. Takeda, T. Otani, K. Okada, H. Onodera, M. Ugaki, Y. Tanaka, C. Tomiyama-Akimoto, T. Yamaguchi, E. Minami, N. Shibuya, T. Omori, M. Nishiyama, H. Nojiri, and H. Yamane. 2004. RERJ1, a jasmonic acid-responsive gene from rice, encodes a basic helix-loop-helix protein. Biochem. Biophys. Res. Commun. 325(3): 857–863.

Klukas, C., D. Chen, and J.-M. Pape. 2014. Integrated Analysis Platform: An Open-Source Information System for High-Throughput Plant Phenotyping. Plant Physiol. 165(2): 506–518Available at http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4044849&tool=pmcentrez&rendertype=abstract.

Kvilekval, K., D. Fedorov, B. Obara, A. Singh, and B.S. Manjunath. 2009. Bisque: A platform for bioimage analysis and management. Bioinformatics 26(4): 544–552.

Lenth, R. V, and M. Hervé. 2015. lsmeans: Least-Squares Means.

Lin, S.-L., F.-L. Chang, S.-Y. Ho, P. Charoenkwan, K.-W. Wang, and H.-L. Huang. 2015. Predicting Neuroinflammation in Morphine Tolerance for Tolerance Therapy from Immunostaining Images of Rat Spinal Cord. PLoS One 10(10): e0139806.

Müller-Linow, M., F. Pinto-Espinosa, H. Scharr, and U. Rascher. 2015. The leaf angle distribution of natural plant populations: assessing the canopy with a novel software tool. Plant Methods 11(1): 11Available at

http://www.plantmethods.com/content/11/1/11.

Nakagawa, M., K. Shimamoto, and J. Kyozuka. 2002. Overexpression of RCN1 and RCN2, rice TERMINAL FLOWER 1 CENTRORADIALIS homologs, confers delay of phase. 29.

Nakamura, A., I. Umemura, K. Gomi, Y. Hasegawa, H. Kitano, T. Sazuka, and M. Matsuoka. 2006. Production and characterization of auxin-insensitive rice by overexpression of a mutagenized rice IAA protein. Plant J. 46(2): 297–306.

Ortiz, A., J.M. Gorriz, J. Ramírez, D. Salas-Gonzalez, A.D.N. Initiative, and others. 2013. Improving MRI segmentation with probabilistic GHSOM and multiobjective optimization. Neurocomputing 114: 118–131.

Peng, S., G.S. Khush, P. Virk, Q. Tang, and Y. Zou. 2008. Progress in ideotype breeding to increase rice yield potential. F. Crop. Res. 108(1): 32–38.

Pordes, R., D. Petravick, B. Kramer, D. Olson, M. Livny, A. Roy, P. Avery, K. Blackburn, T. Wenaus, F. Würthwein, and others. 2007. The open science grid. p. 12057. *In* Journal of Physics: Conference Series.

Pound, M.P., A.P. French, J.A. Atkinson, D.M. Wells, M.J. Bennett, and T. Pridmore. 2013. RootNav: Navigating Images of Complex Root Architectures. Plant Physiol. 162(4): 1802–1814Available at http://www.plantphysiol.org/cgi/doi/10.1104/pp.113.221531.

R Core Team. 2014. R: A Language and Environment for Statistical Computing.

Rojo-Arreola, L., T. Long, D. Asarnow, B.M. Suzuki, R. Singh, and C.R. Caffrey. 2014.

Chemical and genetic validation of the statin drug target to treat the helminth disease, schistosomiasis. PLoS One 9(1).

Shannon, S., and D. Meeks-Wagner. 1991. A Mutation in the Arabidopsis TFL1 Gene Affects Inflorescence Meristem Development. Plant Cell 3(September): 877–892.

Sozzani, R., and P.N. Benfey. 2011. High-throughput phenotyping of multicellular organisms: finding the link between genotype and phenotype. Genome Biol. 12(3): 219Available at http://www.ncbi.nlm.nih.gov/pubmed/21457493%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3129668.

Tanabata, T., T. Shibaya, K. Hori, K. Ebana, and M. Yano. 2012. SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. Plant Physiol. 160(December): 1871–1880.

Tandon, J.P., and H.K. Jain. 2004. Plant ideotype: the concept and application. p. 585–600. *In* Plant Breeding. Springer.

Thain, D., T. Tannenbaum, and M. Livny. 2005. Distributed computing in practice: The Condor experience. Concurr. Comput. Pract. Exp. 17(2–4): 323–356.

Whan, A.P., A.B. Smith, C.R. Cavanagh, J.-P.F. Ral, L.M. Shaw, C.A. Howitt, and L. Bischof. 2014. GrainScan: a low cost, fast method for grain size and colour measurements. Plant Methods 10(1): 23.

Wickland, D.P., and Y. Hanzawa. 2015. The FLOWERING LOCUS T/TERMINAL FLOWER 1 Gene Family: Functional Evolution and Molecular Mechanisms. Mol.

Plant 8(7): 983–997.

Yamamoto, T., F. Taguchi-Shiobara, Y. Ukai, T. Sasaki, and M. Yano. 2001. Mapping Quantitative Trait Loci for Days-to-heading, and Culm, Panicle and Internode Lengths in a BC1F3 Population Using an Elite Rice Variety, Koshihikari, as the Recurrent Parent. Breed. Sci. 51(2): 63–71.

Yang, W., L. Duan, G. Chen, L. Xiong, and Q. Liu. 2013. Plant phenomics and high-throughput phenotyping: Accelerating rice functional genomics using multidisciplinary technologies. Curr. Opin. Plant Biol. 16(2): 180–187Available at http://dx.doi.org/10.1016/j.pbi.2013.03.005.

Yazdanbakhsh, N., and J. Fisahn. 2009. High throughput phenotyping of root growth dynamics, lateral root formation, root architecture and root hair development enabled by PlaRoM. Funct. Plant Biol. 36(11): 938–946.

Zhao, K., C.-W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2(1–4): 467.

3.7 Figures

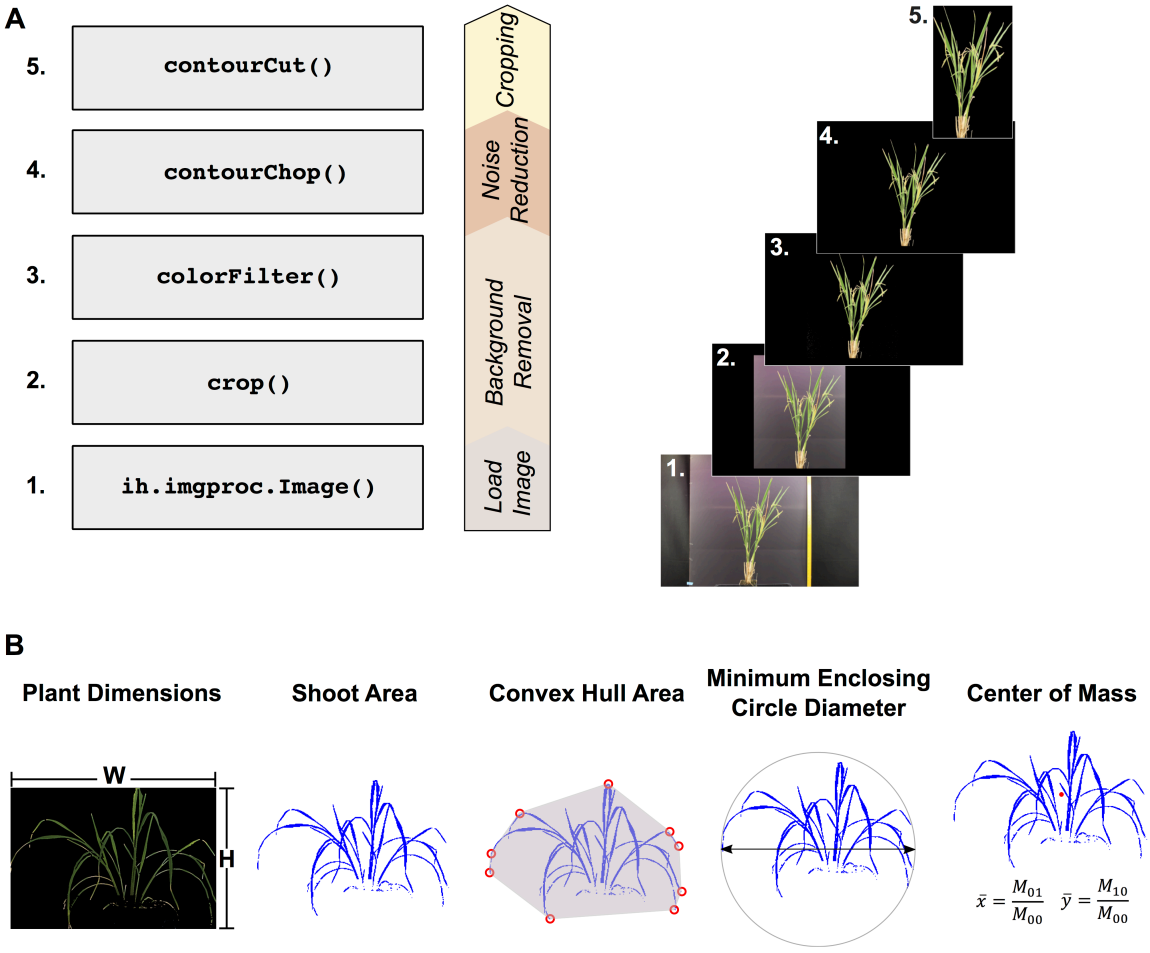


Figure 1. Simple processing workflow for side view RGB images and digital morphological descriptors. (A) Major processing tasks used to process RGB side view images of rice plants captured with a conventional SLR camera. The full processing script, which lists all processing tasks and the appropriate parameters is provided as Supplemental File S1. The panels on the right show the results of each processing step. (B) A subset of the morphological digital traits currently available in IH. A complete list of the digital traits used is provided in Supplemental Table S3, and are visually depicted in Supplemental Fig. S3. The user-defined region of interest used for height calculations

is indicated with the red box in Plant Dimensions. The center of mass is indicated by the red point in the center of the plant.

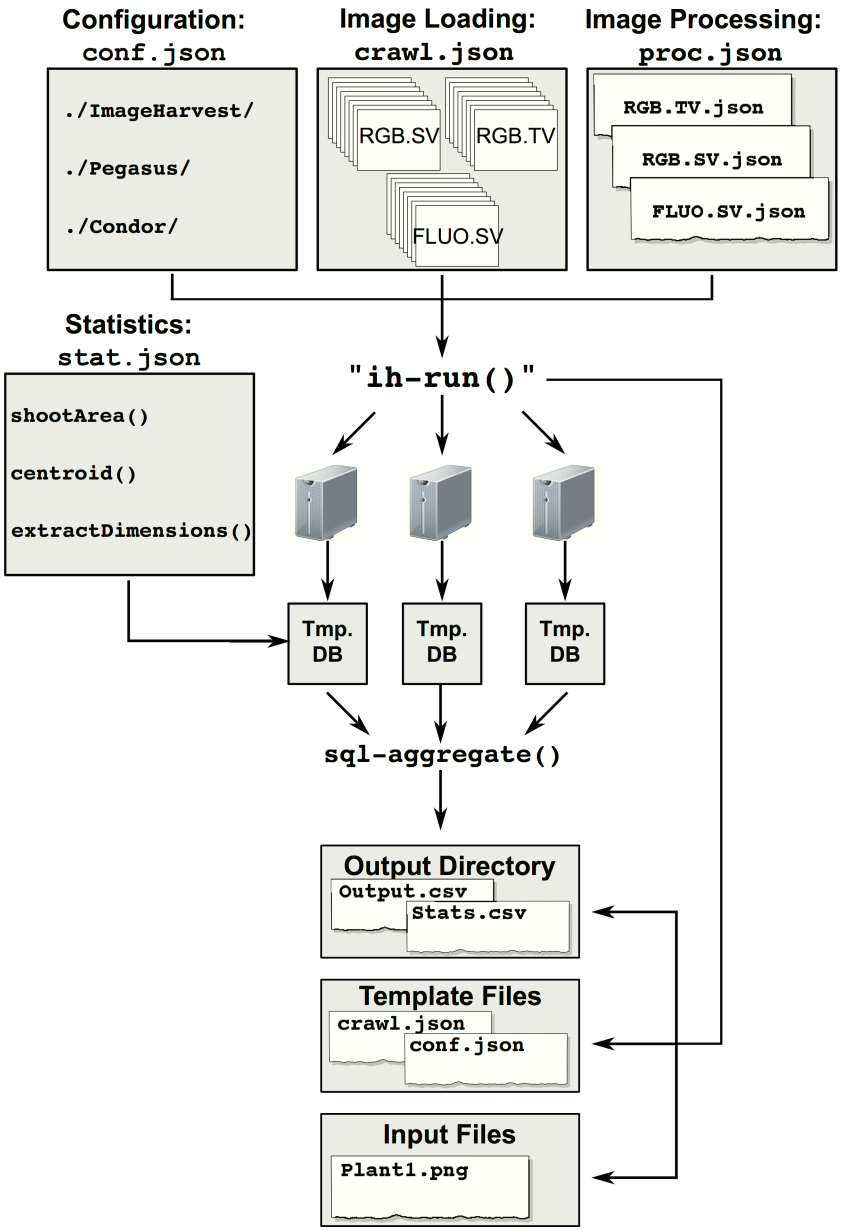


Figure 2. Schematic summarising the creation and execution of supercomputing workflows using IH. IH requires four types of template files that contain the necessary information for image loading, processing, system configuration and statistics workflows. The execution of the "`ih-run`" command creates three directories, which contain the

final output database, a copy of the submitted template files, and all non-image raw-input files. The "`sql-aggregate`" function collects digital metric data from the intermediate databases, compiles the data and writes it to the output directory.

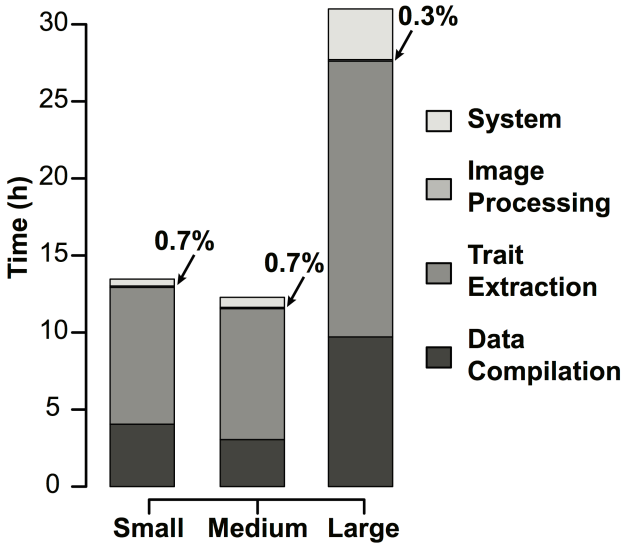


Figure 3. Computational time for image sets of various sizes. The processing speed of RGB and FLUO workflows was evaluated using three image sets of various size: small (19,475 images), medium (38,357 images) and large (77,665 images). The tasks were separated into four categories: system (Pegasus and condor related tasks), processing (actual image processing functions), trait extraction ("`ih-extract-multi`", "`ih-stats-histogram`") and data compilation tasks ("`ih-sql-aggregate`"). Time refers to the time experienced on the user's end.

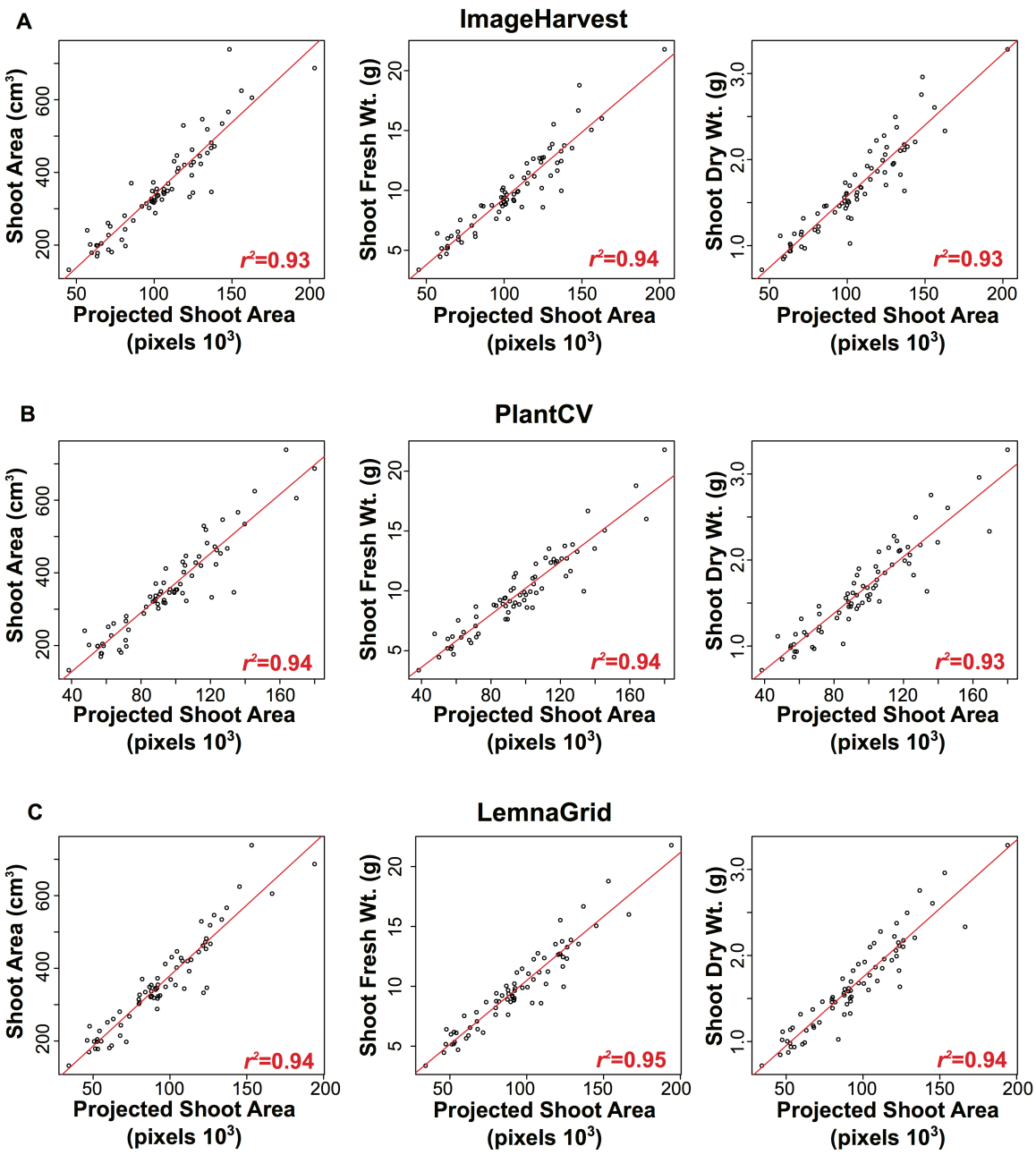


Figure 4. Comparisons between projected shoot area derived from several publically available plant image processing software and conventional phenotypic measurements of plant biomass. Projected shoot area, here defined as the summation of plant pixels from two side view images, was calculated using (A) IH, (B) PlantCV and (C) LemnaGrid and compared to three manual measurements of biomass. Manual

phenotypic measurements were recorded on 72 rice plants 28 days after transplanting. Correlation analysis was performed using Pearson's method. All correlations were significant ($p < 0.0001$).

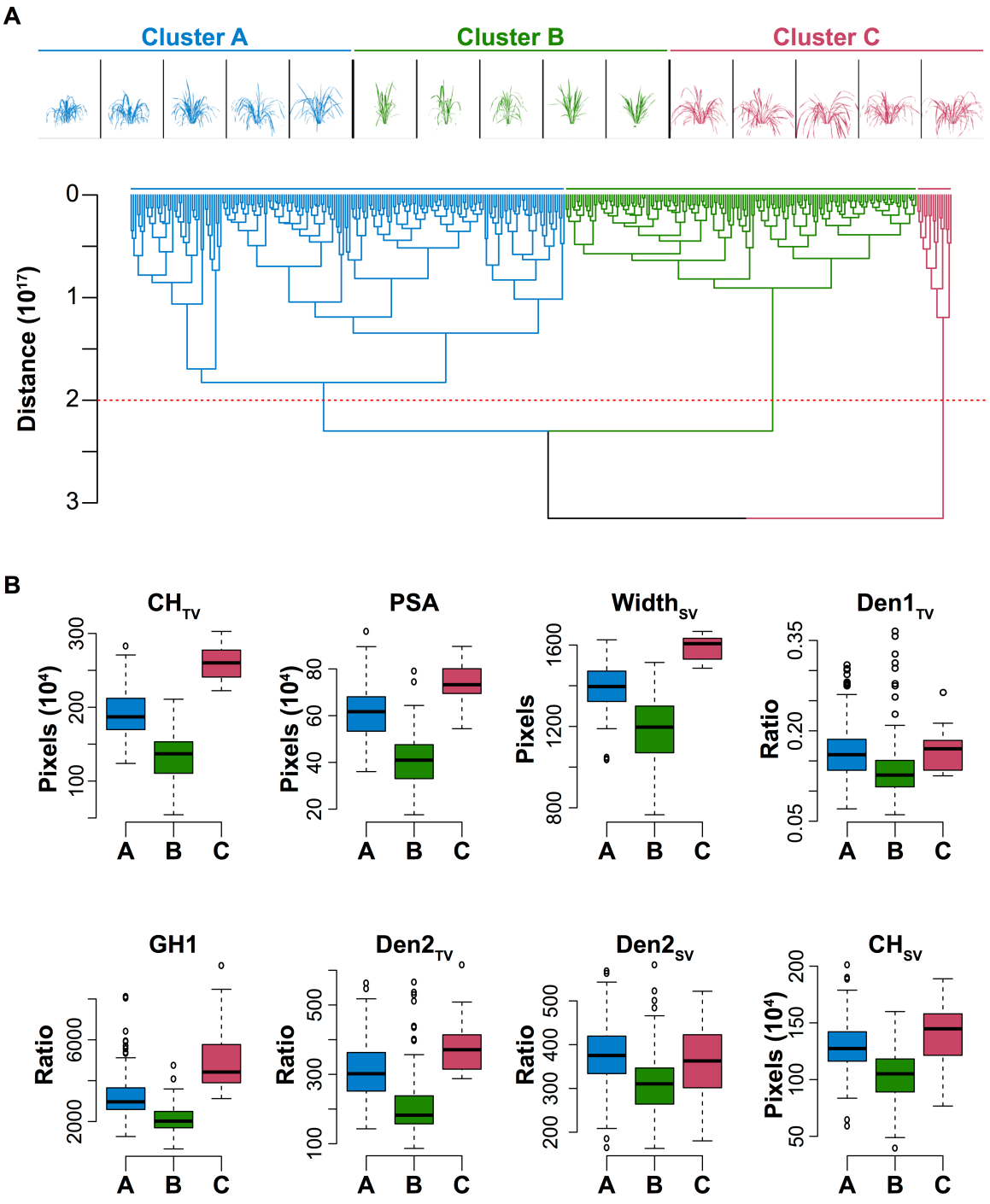


Figure 5. Hierarchical clustering of raw image moments (M) and the distribution of a subset of digital traits for each cluster. (A) Hierarchical clustering of raw image moments extracted from both sideview and top view images of rice plants at 41 days after transplant. Five representative plants from each of the three clusters are shown above the dendrogram. A Euclidian distance threshold of 2 $x10^{17}$ was used to separate the clusters, and is indicated by the horizontal red dashed line. (B) Boxplots for eight of the 22 digital traits illustrating the phenotypic distributions of the major clusters. Boxplots for the remaining 14 digital traits is provided as Supplemental Fig. S2. CH: Convex hull; GH: growth habit; Den: density; PSA: projected shoot area; TV: top view; SV: side-view

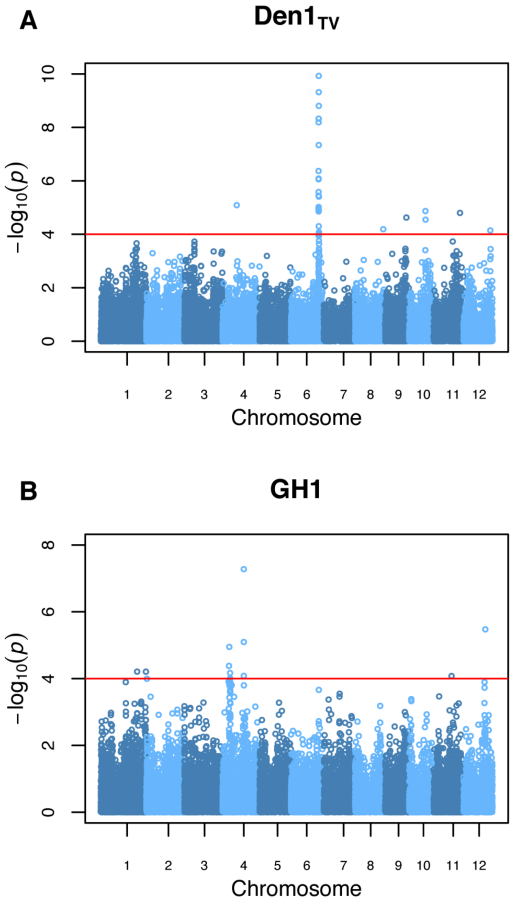


Figure 6. Genome wide association analysis for $Den1_{TV}$ (A) and GH1 (B). A threshold of $p < 10^{-4}$ was used to identify significant associations and is indicated by the red horizontal line in each plot. Den: Density; GH: Growth Habit; TV: Top View; SV: Side View

3.8 Tables

Table 1. Candidate genes and QTLs overlapping with other studies. *GH*: Growth Habit; *Den*: Density; *PSA*: Projected Shoot Area; *CA*: Cropped Area; Chr.: Chromosome

QTL Position (bp)	Trait(s)	Candidate Gene/ Reported QTLs	Reference
20,144,020-20,622,471	GH1, GH2, GH3, GH4	*OsTFL1*	(Nakagawa et al., 2002)
13,268,374-13,668,374	$Den1_{TV}$, $Den3_{TV}$, $Den2_{SV}$, PSA	*OsRERJ1*	(Kiribuchi et al., 2004)
26,031,137-26,892,628	$Den1_{TV}$, $Den2_{TV}$, $Den3_{TV}$, $Den1_{SV}$, $Den3_{SV}$, CA_{SV}, $Width_{SV}$	*qIN3-6*	(Yamamoto et al., 2001)
		qPH6-2	(Duan et al., 2013)
25,302,716-25,702,716	$Den1_{TV}$, $Den3_{TV}$	*OsIAA3*	(Nakamura et al., 2006)

3.9 Supplemental Materials

Figure S1. Imaging environment used for phenotyping with a conventional SLR camera. An imaging room was constructed to image a rice plant (cv 9311) at maturity in a homemade. To improve image processing several lighting sources were used to provide adequate light and a uniform background.

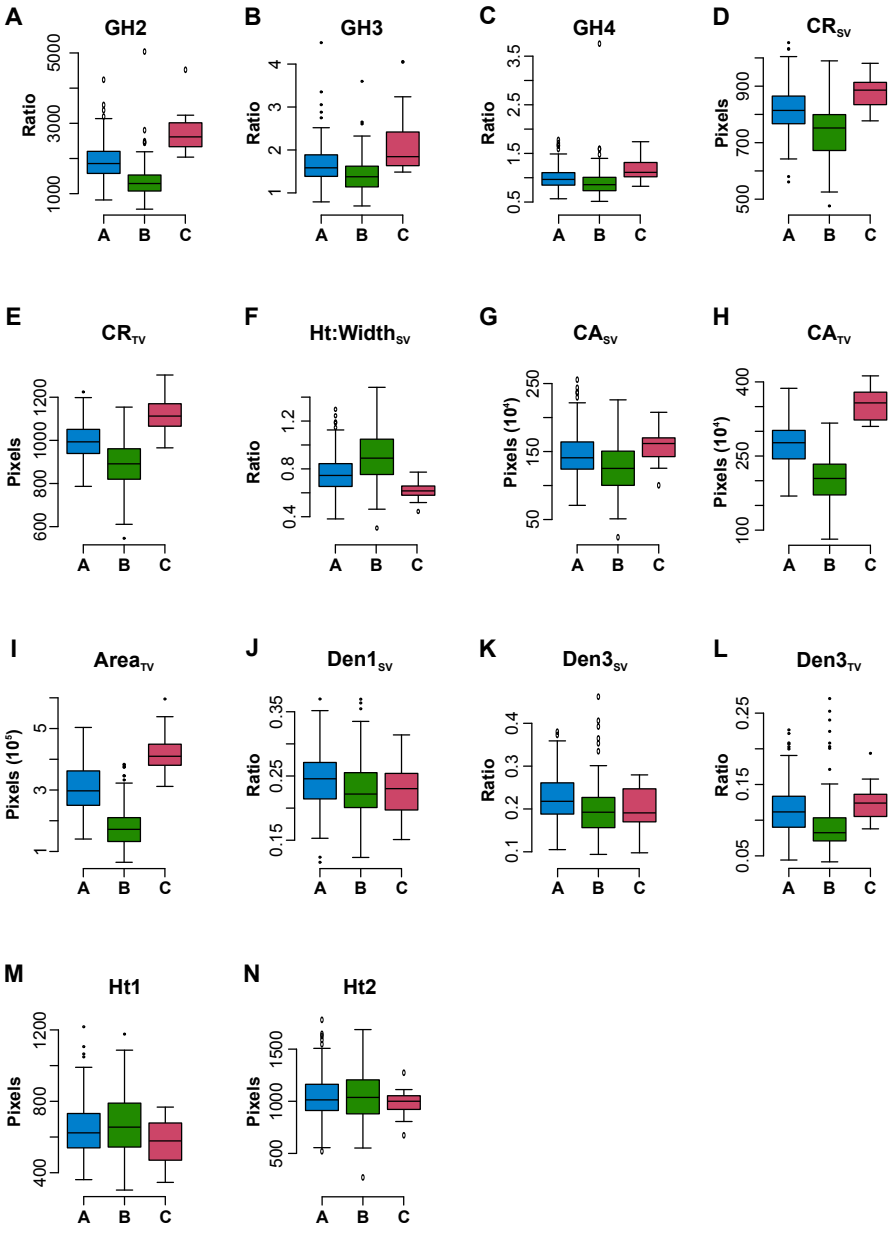


Figure S2. Boxplots summarizing the phenotypic distribution within each cluster for 14 digital traits (A-N). Hierarchical clustering of raw image moments identified three major phenotypic groups indicated by the letters below each bar. A full description of each of the digital traits, as well as the formula used for calculations is provided in Table S2. *Ht*: Height; *CH*: Convex Hull Area; *CR*: Radius of minimum enclosing circle; *Den*: Density; *GH*: Growth habit

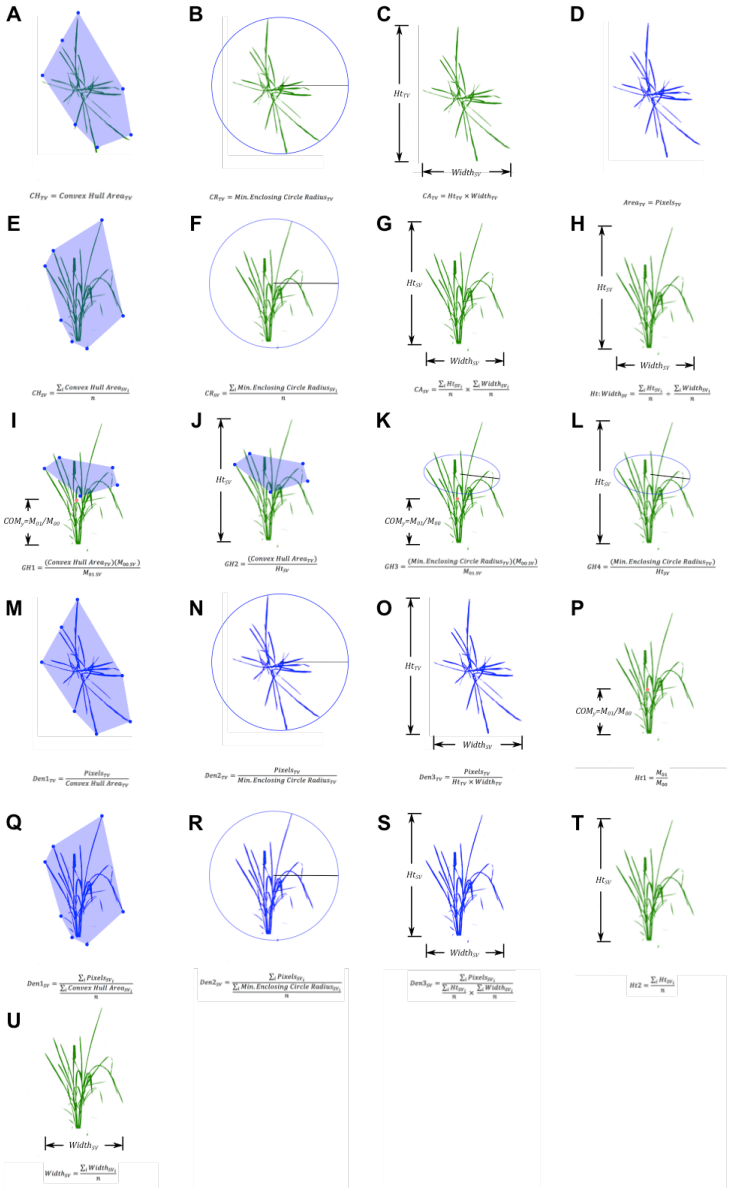


Figure S3. Visualization of 21 of the 22 digital traits derived from IH metrics (A-U). Each trait is visualized using a representative plant, and the equations used for trait calculations are listed below each panel. The summation of plant pixels from one or more imaging perspective was used in the calculation of several digital traits and is indicated by plant contours filled in blue. Convex hull area is indicated with the light blue shaded bounding polygon superimposed over portions of the plant. COM_y: Center of mass about the *y*-axis; *Ht*: Height; *CH*: Convex Hull Area; *CR*: Radius of minimum enclosing circle; *Den*: Density; *GH*: Growth Habit; *M*: raw moment; *n*: number of side view imaging perspectives.

Table S1. A complete list of functions that are available in Image Harvest.

Function	Function Class	Function Sub-class	Brief Description
destroy()	ih.imgproc.Image method	Base	Destroys all currently open windows.
list()	ih.imgproc.Image method	Base	Lists all saved states.
resize()	ih.imgproc.Image method	Base	Resize an image.
restore()	ih.imgproc.Image method	Base	Reloads a previously saved image from the 'states' variable.
save()	ih.imgproc.Image method	Base	This function saves the current image in the 'states' variable under the specified name. It can then be reloaded using the restore() method.
show()	ih.imgproc.Image method	Base	Displays the image in a window.
wait()	ih.imgproc.Image method	Base	Waits until a key is pressed, then destroys all windows and continues program execution.
write()	ih.imgproc.Image method	Base	Writes the current image to the given output directory, with the given name.
anova()	ih.statistics.Stats method	Data Analysis	Computes the analysis of variation of all numeric information based on 3 factors, treatment, date, and the interaction between the two. Analysis of variation is different than the rest of the stats functions, in that a lot of information is lost after running it. The results themselves correspond to columns (pixels, rmed, binx...) instead of actual images.
correlation()	ih.statistics.Stats method	Data Analysis	This function correlates all numeric values with values in the given file. The input data file is assumed to be in csv format.
logErrors()	ih.statistics.Stats method	Data Analysis	This function writes all errors from a given table to a log file, usually used at the end of image processing to write all images that did not process correctly.
export()	ih.statistics.Stats method	Data Analysis	This function simply extracts data from a database and writes it to csv format. Default functionality is to extract only data that has been processed.

Table S1 (con't). A complete list of functions that are available in Image Harvest.

Function	Function Class	Function Sub-class	Brief Description
extractPixels()	ih.imgproc.Image method	Post-process	Returns the number of non-black (foreground) pixels in the image. Creates a temporary binary image to do this.
normalize()	ih.statistics.Stats method	Data Analysis	Normalizes all numerical information to the specific column. This function is usually used with 'pixels' as the specified column, which expresses all numeric information as a percentage of the total pixels in the image.
shootArea()	ih.statistics.Stats method	Data Analysis	This function sums the numeric values of multiple image types together. In general, it is used to combine side view + top view images of the same spectrum.
treatmentComp()	ih.statistics.Stats method	Data Analysis	This function compares information between treatments – It finds plants that are identical except for treatment, and computes either a ratio or difference between them.
tTest()	ih.statistics.Stats method	Data Analysis	This function computes a ttest of the input table for all numeric headers.
extractBins()	ih.imgproc.Image method	Post-process	This function counts the number of pixels that fall into the range as specified by each bin.
extractColorData()	ih.imgproc.Image method	Post-process	This function calculates a normalized histogram of each individual color channel of the image, and returns the mean & median of the histograms for the channels specified.
extractConvexHull()	ih.imgproc.Image method	Post-process	Returns the area of the convex hull around all non black pixels in the image.
extractDimensions()	ih.imgproc.Image method	Post-process	Returns a list corresponding to the height and width of the image.
extractMinEnclosingCircle()	ih.imgproc.image method	Post-process	Returns the center and radius of the minimum enclosing circle of all non-black pixels in the image.
extractMoments()	ih.imgproc.Image method	Post-process	Calculates the moments of the image, and returns a dicitonary based on them. This function is a wrapper to the OpenCV function moments.
extractFinalPath()	ih.imgproc.Image method	Post-process	This function writes the absolute path of the output file to the database.

Table S1 (con't). A complete list of functions that are available in Image Harvest.

Function	Function Class	Function Sub-class	Brief Description
mask()	ih.imgproc.Image method	Processing	This function converts the image to a mask by performing convertColor("bgr", "gray"), convertColor("gray", "bgr")
adaptiveThreshold()	ih.imgproc.Image method	Processing	Thresholds an image by considering the image in several different windows instead of the image as a whole. This function is a wrapper to the OpenCV function adaptiveThreshold.
bitwise_and()	ih.imgproc.Image method	Processing	Performs logical AND between the input image and the comp image.
bitwise_not()	ih.imgproc.Image method	Processing	Inverts the image. If the given image has multiple channels (i.e. is a color image) each channel is processed independently.
bitwise_or()	ih.imgproc.Image method	Processing	Performs logical OR between the input image and the comp image.
bitwise_xor()	ih.imgproc.Image method	Processing	Performs exclusive logical OR between the input image and the comp image.
blur()	ih.imgproc.Image method	Processing	Smoothes an image using the normalized box filter. This function is a wrapper to the OpenCV function blur.
colorFilter()	ih.imgproc.Image method	Processing	This function applies a color filter defined by the input logic, to a targeted region defined by the input region of interest (ROI).
contourCut()	ih.imgproc.Image method	Processing	This function crops an image based on the size of detected contours in the image – clusters of pixels in the image. The image is cropped such that all contours that are greater than the specified area are included in the final output image.
crop()	ih.imgproc.Image method	Processing	This function crops the image based on the given region of interest (ROI) [ystart, yend, xstart, xend].
edges()	ih.imgproc.Image method	Processing	This function calculates the edges of an image using the Canny edge detection algorithm. This function is a wrapper to the OpenCV function Canny.
gaussianBlur()	ih.imgproc.Image method	Processing	This function blurs an image based on a Gaussian kernel.
convertColor()	ih.imgproc.Image method	Processing	Converts the given image between color spaces, based on the given types.

Table S1 (con't). A complete list of functions that are available in Image Harvest.

Function	Function Class	Function Sub-class	Brief Description
kmeans()	ih.imgproc.Image method	Processing	This function is a wrapper to the OpenCV function kmeans Reduced the number colors in the image to the most compact 'central' colors (k). The number of colors in the resulting image is the specified value 'k'.
knn()	ih.imgproc.Image method	Processing	Classifies and removes pixels based on the nearest neighbors algorithm. This function is a wrapper to the OpenCV function KNearest.
meanshift()	ih.imgproc.Image method	Processing	Segments the image into clusters based on nearest neighbors. This function is a wrapper to the pymeanshift module.
medianBlur()	ih.imgproc.Image method	Processing	This function smooths an image using the median filter.
morphology()	ih.imgproc.Image method	Processing	This function performs morphological operations based on the inputted values. This function is a wrapper to the OpenCv function morphologyEx.
normalizeByIntensity()	ih.imgproc.Image method	Processing	Normalizes each channel of the pixel by its intensity.
resizeSelf()	ih.imgproc.Image method	Processing	Resizes the current image.
threshold()	ih.imgproc.Image method	Processing	Thresholds the image based on the given type. The image must be grayscale to be thresholded.
dimFromROI()	ih.imgproc.Image method	Post-process	Returns a list corresponding to the height (defined as the distance from the upper most plant pixel to a upper boundary of a user-defined region of interest (ROI)) and width of the image.

Table S2. Arguments and metrics returned from `"ih-extract-multi"`.

Arguments	**Description**
--pixels	Counts the number of plant pixels in the image.
--colors	Calculates a normalized histogram of each individual color channel of the image, and returns the mean & median of the histograms for the channels specified.
--channels	This function extracts the total number of pixels of each color value (0 to 255) for each channel.
--moments	Calculates the moments of the image, and returns a dicitonary based on them. Spatial moments are prefixed with 'm', central moments are prefixed with 'mu', and central normalized moments are prefixed with 'nu'. This function is a wrapper to the OpenCV function moments.
--hull	Calculates the area of the convex hull around all non black pixels in the image.
--circle	Calculates the center and radius of the minimum enclosing circle encompasing all non-black pixels in the image.
--dimensions	Calculates the height and width of the image.
--bins	Counts the number of pixels that fall into an RGB range. This function accepts a list of ranges, which is defined by six values (three minimum RGB and three maximum RGB values).

Table S3. Digital traits used to describe plant morphological qualities. SV: side view; TV: top view; *n*: number of side view images

Digital Trait	Class	Formula	Description
$Area_{TV}$	Biomass	$Area_{TV} = Pixels_{TV}$	Plant pixels from extracted from top view (TV) image
PSA	Biomass	$PSA = Pixels_{TV} + \sum_i Pixels_{SV_i}$	Summation of the plant pixels from all side view images and top view
$Ht{:}Width_{SV}$	Plant Shape	$Ht{:}Width_{SV} = \frac{\sum_i Ht_{SV_i}}{n} \div \frac{\sum_i Width_{SV_i}}{n}$	Ratio of the average height of the cropped SV image to the average width of the cropped SV image with dimFromROI()
CA_{TV}	Plant Size	$CA_{TV} = Ht_{TV} \times Width_{TV}$	Product of the height and width of the cropped TV image with dimFromROI()
$Den1_{SV}$	Density/ Compactness	$Den1_{SV} = \frac{\sum_i Pixels_{SV_i}}{\frac{\sum_i Convex\ Hull\ Area_{SV_i}}{n}}$	Ratio of the summation of the plant pixels from all SV images and the average convex hull area of all SV images
$Den2_{SV}$	Density/ Compactness	$Den2_{SV} = \frac{\sum_i Pixels_{SV_i}}{\frac{\sum_i Min.Enclosing\ Circle\ Radius_{SV_i}}{n}}$	Summation of the plant pixels from all SV images to the average radius of the minimum enclosing circle of all SV images
$Den3_{SV}$	Density/ Compactness	$Den3_{SV} = \frac{\sum_i Pixels_{SV_i}}{\frac{\sum_i Ht_{SV_i}}{n} \times \frac{\sum_i Width_{SV_i}}{n}}$	Summation of the plant pixels from all SV images to the average side view cropped area (CA_{SV}) with dimFromROI()
$Den1_{TV}$	Density/ Compactness	$Den1_{TV} = \frac{Pixels_{TV}}{Convex\ Hull\ Area_{TV}}$	Ratio of the plant pixels extracted from TV image to the convex hull area of the TV image
$Den2_{TV}$	Density/ Compactness	$Den2_{TV} = \frac{Pixels_{TV}}{Min.Enclosing\ Circle\ Radius_{TV}}$	Ratio of the plant pixels extracted from TV image to the radius of the minimum enclosing circle of the TV image
$Den3_{TV}$	Density/ Compactness	$Den3_{TV} = \frac{Pixels_{TV}}{Ht_{TV} \times Width_{TV}}$	Ratio of the plant pixels extracted from TV image to the top view cropped area (CA_{TV})
Ht2	Height	$Ht2 = \frac{\sum_i Ht_{SV_i}}{n}$	Average height of the final cropped image from all side views determined with dimFromROI()
Ht1	Height	$Ht1 = \frac{M_{01}}{M_{00}}$	Center of mass about the *y*-axis determined from the raw image moments (*M*) of side view images
GH1	Plant Shape	$GH1 = \frac{(Convex\ Hull\ Area_{TV})(M_{00\ SV})}{M_{01\ SV}}$	The ratio of the convex hull area of the TV image to the center of mass about the *y*-axis of all side view images
GH2	Plant Shape	$GH2 = \frac{(Convex\ Hull\ Area_{TV})}{Ht_{SV}}$	The ratio of the convex hull area of the TV image to the average height of all final cropped SV images determined with dimFromROI()

Table S3 (*cont'd*). Digital traits used to describe plant morphological qualities.

Digital Trait	Class	Formula	Description
GH3	Plant Shape	$GH3 = \frac{(Min.\,Enclosing\ Circle\ Radius_{TV})(M_{00\,SV})}{M_{01\,SV}}$	The ratio of the minimum enclosing circle radius of the TV image to the average center of mass about the y-axis of all SV images
GH4	Plant Shape	$GH4 = \frac{(Min.\,Enclosing\ Circle\ Radius_{TV})}{Ht_{SV}}$	The ratio of the minimum enclosing circle radius of the TV image to the average height all final cropped SV images determined with dimFromROI()
$Width_{SV}$	Plant Size	$Width_{SV} = \frac{\sum_i Width_{SV_i}}{n}$	The average width of all final cropped SV images
CA_{SV}	Plant Size	$CA_{SV} = \frac{\sum_i Ht_{SV_i}}{n} \times \frac{\sum_i Width_{SV_i}}{n}$	Product of the average height and width of all cropped SV images determined with dimFromROI()
CH_{TV}	Plant Size	$CH_{TV} = Convex\ Hull\ Area_{TV}$	Convex hull area of the TV image
CH_{SV}	Plant Size	$CH_{SV} = \frac{\sum_i Convex\ Hull\ Area_{SV_i}}{n}$	Average convex hull area of all SV images
CR_{TV}	Plant Size	$CR_{TV} = Min.\,Enclosing\ Circle\ Radius_{TV}$	The radius of the minimum enclosing circle radius of the TV image
CR_{SV}	Plant Size	$CR_{SV} = \frac{\sum_i Min.\,Enclosing\ Circle\ Radius_{SV_i}}{n}$	The average radius of the minimum enclosing circle radius of all SV images

Table S4. ANOVA results and phenotypic means for each digital trait and cluster. A one-way ANOVA was conducted to compare digital traits between each cluster. A threshold of $p < 0.0001$ was used to determine statistical significance.

Digital Trait	Cluster A		Cluster B		Cluster C		F-Value	p-Value
	Mean	SEM	Mean	SEM	Mean	SEM		
Ht2	1044.4	15.62	1053.6	18.97	984.4	34.50	0.684	0.5052
$Width_{SV}$	1391.3	7.63	1183.9	11.95	1586.2	14.60	157.799	2.2E-50
CA_{SV}	1456183.8	24100.20	1262268.2	28338.38	1564989.8	63311.77	16.467	1.40E-07
$Ht:Width_{SV}$	0.757	0.0117	0.904	0.0167	0.617	0.0191	37.579	1.35E-15
CA_{TV}	2757517.7	29851.35	2034003.6	37109.33	3544363.7	77959.15	169.499	4.33E-53
$Den1_{SV}$	0.242	0.0031	0.229	0.0037	0.225	0.0109	4.750	0.0092
$Den2_{SV}$	378.7	5.2217	313.6	5.5896	358.8	22.8188	35.224	9.72E-15
$Den3_{SV}$	0.222	0.0037	0.199	0.0047	0.203	0.0126	7.991	4.00E-04
$Area_{TV}$	306706.4	5372.40	178276.7	4972.82	425004.9	17803.17	195.642	7.88E-59
PSA	617370.6	7861.68	413742.7	8444.76	741479.0	23130.37	186.447	7.4E-57
$Den1_{TV}$	0.166	0.0033	0.137	0.0040	0.169	0.0090	16.637	1.20E-07
$Den2_{TV}$	312.7	6.08	204.5	6.49	385.1	21.00	89.241	2.13E-32
$Den3_{TV}$	0.116	0.0024	0.092	0.0030	0.124	0.0068	21.051	2.18E-09
Ht1	648.0	11.15	675.1	13.33	568.0	32.23	3.752	0.0244
CH_{TV}	1912150.1	22566.88	1343149.6	24169.27	2601697.5	61207.56	219.917	8.1E-64
CH_{SV}	1295565.6	16081.97	1039990.1	17641.92	1388537.5	72797.46	61.956	5.85E-24
GH1	3239.3	76.46	2101.9	51.31	5171.0	484.51	108.566	6.94E-38
GH2	1930.0	37.14	1356.9	38.07	2719.4	149.99	88.371	3.84E-32
CR_{TV}	992.2	5.81	887.0	9.04	1120.4	19.47	76.954	1.05E-28
CR_{SV}	820.5	5.47	747.5	7.63	879.6	13.91	41.195	6.84E-17
GH3	1.674	0.0341	1.409	0.0302	2.217	0.2116	30.140	7.35E-13
GH4	1.001	0.0157	0.905	0.0241	1.166	0.0541	10.822	2.70E-05

Supplemental File S1. Candidate genes within a 200kb window of significant SNPs. A window of 200kb was chosen based on the estimated linkage disequilibrium in rice (Zhao et al., 2011).

CHAPTER 4

A COMPREHENSIVE IMAGE-BASED PHENOMICS ANALYSIS REVEALS THE COMPLEX GENETIC ARCHITECTURE OF SHOOT GROWTH DYNAMICS IN RICE (*ORYZA SATIVA*)

Malachy T Campbell, Qian Du, Kan Liu, Chris J Brien, Bettina Berger, Chi Zhang, Harkamal Walia

Published: November 22, 2016; doi: 10.3835/plantgenome2016.07.0064

Abstract

Early vigor is an important trait for many rice-growing environments. However, genetic characterization and improvement for early vigor is hindered by the temporal nature of the trait and high genotype-by-environment effects. We explored the genetic architecture of shoot growth dynamics during the early and active tillering stages by applying a functional modeling and genome-wide association (GWAS) mapping approach on a diversity panel of ~360 rice accessions. Multiple loci with small effects on shoot growth trajectory were identified, indicating a complex polygenic architecture. Natural variation for shoot growth dynamics was assessed in a subset of 31 accessions using RNA sequencing and hormone quantification. These analyses yielded a GA catabolic gene, *OsGA2ox7*, which could be influencing GA levels to regulate vigor in the early tillering stage. Given the complex genetic architecture of shoot growth dynamics, the potential of genomic selection (GS) for improving early vigor was explored using all 36,901 SNPs as well as several subsets of the most significant SNPs from GWAS. Shoot growth trajectories could be predicted with reasonable accuracy using the top 50 most significant SNPs from GWAS (0.37-0.53), however the accuracy of prediction improved by including more

markers, which indicates that GS may be an effective strategy for improving shoot growth dynamics during the vegetative growth stage. This study provides insights into the complex genetic architecture and molecular mechanisms underlying early shoot growth dynamics, and provides a foundation for improving this complex trait in rice.

4.1 Introduction

Early vigor, defined, as the plant's ability to rapidly accumulate shoot biomass during early developmental stages, is critical for stand establishment, resource acquisition and ultimately yield. The rapid emergence of leaves leads to early canopy closure, which reduces soil evaporation, thereby improving seasonal water use efficiency and conserving water for later vegetative growth and grain production. In rice, early vigor is a particularly important trait for regions where rice is direct seeded (Mahender et al., 2015). As the cost of labor rises, a shift from the labor-intensive practice of transplanted rice to direct-seeded rice is the expected solution to solve this problem (Mahender et al., 2015).

Several studies have examined seedling vigor in rice and elucidated the underlying genetic basis using conventional phenotyping strategies in field and greenhouse conditions (Redoña and Mackill, 1996; Lu et al., 2007; Cairns et al., 2009; Rebolledo et al., 2012a; b, 2015; Liu et al., 2014). In a recent study by Rebolledo *et al* (2015), multiple vigor-related traits such as plant morphology and non-structural carbohydrates were quantified in a rice diversity panel of 123 *japonica* varieties (Rebolledo et al., 2015). The authors integrated multiple phenotypic metrics in a functional–structural plant model, called Ecomeristem, and performed genome-wide association mapping using phenotypic metrics and model parameters as trait values (Luquet et al., 2012; Rebolledo et al., 2015). Such multi-trait approaches provide a more comprehensive

understanding of the biochemical and genetic basis of early vigor than conventional single trait approaches.

Early vigor is a function of time. The timing of developmental switches that initiate tiller formation and rapid exponential growth are a crucial component of this trait. However, despite this temporal dimension, most studies have assessed the genetic basis of early vigor at one or a few discrete time points (Redoña and Mackill, 1996; Lu et al., 2007; Cairns et al., 2009; Rebolledo et al., 2012a; b, 2015; Liu et al., 2014). Such approaches are overly simplistic and may only provide a snapshot of the genetic determinants that cumulatively influence the final biomass. However, sampling for biomass at high frequencies over a developmental window for mapping populations using conventional destructive phenotyping approaches would require tens to hundreds of thousands of plants and is highly labor intensive. With the advent of high-throughput, image-based phenomics platforms, plants can be phenotyped non-destructively more frequently throughout the growth cycle to examine the temporal dynamics of physiological and morphological traits (Berger et al., 2010; Golzarian et al., 2011; Busemeyer et al., 2013; Topp et al., 2013; Moore et al., 2013; Würschum et al., 2014; Hairmansis et al., 2014; Slovak et al., 2014; Yang et al., 2014; Honsdorf et al., 2014; Chen et al., 2014; Bac-Molenaar et al., 2015).

Mathematical equations that describe a developmental or physiological process can be applied to this high-resolution temporal data to describe temporal growth trajectories using mathematical parameters. Several models, such as logistic, exponential and power-law functions have been used to describe plant growth (Paine et al., 2012). These approaches enhance the temporal resolution of phenotyping and when combined with association or linkage mapping, improves the power of detecting genetic associations for complex traits compared to traditional cross-sectional approaches (Wu and Lin, 2006; Xu et al., 2014; Campbell et al., 2015). However,

despite the recent advances in phenotyping technologies, the genetic basis of early growth dynamics in rice, or other cereals remain largely unexplored.

Multiple and sometimes uncorrelated phenotypes determine the rate and extent of vegetative growth in crops. We hypothesize that capturing growth dynamics at a higher temporal resolution can help elucidate the genetic basis of this trait. To this end, we sought to examine the genetic architecture of temporal shoot growth dynamics during the early and active tillering stages (8-27 days after transplant (DAT) and 19-41 DAT) in rice. A panel of ~360 diverse rice accessions was phenotyped using a non-destructive, image-based platform and temporal trends in shoot growth were modeled with a power law function (Zhao et al., 2011). We provide insights into the genetic basis of shoot growth by using genome wide association (GWAS) analysis. The underlying molecular mechanisms were explored using RNA sequencing on a subset of the diversity panel during the early tillering stage. Genomic selection (GS) of the model parameters and daily estimates of shoot biomass suggest GS may be an effective strategy for improving early vigor in rice.

4.2 Results

4.2.1 Considerable natural variation for shoot growth dynamics is present in Rice Diversity Panel 1

We performed image-based high-throughput phenotyping of a subset of rice diversity panel (RDP1; Famoso et al., 2011; Zhao et al., 2011) consisting of ~360 accessions belonging to each of the five major rice subpopulations (*indica, aus, aromatic, temperate japonica* and *tropical japonica*) with the goal of discovering novel genetic variants influencing temporal growth dynamics during vegetative growth. Two separate, triplicated experiments were

conducted in which the diversity panel was phenotyped over a period of 19 and 22 days at the early vegetative stage (8-27 days after transplant; 361 accessions) and active tillering stage (19-41 days after transplant; 360 accessions), respectively. To quantify growth trajectories as a function of time, the shoot growth dynamics were modeled for each accession using a power-law function (Paine et al., 2012). $M_t = \left({M_o}^{(1-\beta)} + rt(1-\beta)\right)^{1/(1-\beta)}$. Although, rice exhibits a determinate growth pattern, the overlapping developmental windows selected for these studies captures the accelerating phase of shoot growth (Fig. 1 A, B).

Considerable natural variation was observed in the diversity panel for all model parameters (M_o, β and r) as well as discrete measurements for projected shoot area (PSA) (Fig. 2; File S1). Heritability estimates for model parameters ranged from 0.32-0.69 during the early tillering stage, while estimates during the late tillering stage ranged from 0.46-0.72. In both developmental stages, M_o, the initial biomass on the first day of imaging showed the highest heritability (0.69 and 0.72 at the early and active tillering stages, respectively). Model parameters associated with growth rate (β and r) displayed moderate heritability estimates (β: 0.46 and 0.47, and r: 0.32 and 0.46 during the early and active tillering stages, respectively). The moderate to high heritability for all model parameters suggests that a portion of phenotypic variation for growth dynamics is under genetic control. The slightly lower genotypic contribution observed for model parameters describing growth rate compared to M_o indicate a smaller genotypic effect for these rate parameters, and suggest that they may be influenced more by environmental conditions.

To examine the relationship between shoot growth dynamics at the early and active tillering stages, Pearson correlation analysis was conducted for each of the model parameters obtained

from the early tillering and active tillering stages, as well as PSA at each time point in the experiment (File S2). Significant positive relationships were observed for each model parameters between the two developmental stages, suggesting that the growth characteristics observed at early tillering stage partly persistent during the later tillering stages. The model parameter *r* (from both developmental stages) displayed a significant negative correlation with PSA during both the early and later tillering stages. Conversely, β displayed significant positive correlation with PSA, which is not surprising considering that as values of β approach 1, the power-law function begins to behave more like an exponential function, with constant relative growth rates being achieved when β=1. Thus, plants exhibiting near exponential growth tend to be very large.

A temporal trend in correlation was observed between the model parameters r and β and PSA during both developmental stages. For instance, during the early tillering stage β displayed a weak positive relationship with PSA at 8-16 days after transplant (0.25-0.48). However, the strength of the correlation became progressively larger as time progressed with the highest being observed at 23 DAT. The parameter *r* on the other hand, displayed a similar trend with a weak negative relationship observed during the early time points and a progressively stronger negative correlation observed at the later time points. Similar temporal trends were observed for β and *r* during the active tillering stage. These results indicate that parameters obtained from functional modeling (β and *r*) can be used to describe growth characteristics outside of plant size that otherwise would not be identified with conventional cross sectional approaches.

4.2.2 Genetic basis of shoot biomass

To identify QTL that exhibit a time-specific effect on shoot biomass we conducted GWAS at each of the 38 time points during the early tillering and active tillering stages. A total

of seven QTL (26 SNPs) were significantly associated with PSA at one or more time points during the early and active tillering stages ($p < 2.33 \times 10^{-5}$; Table I; File S3). Five QTL (9 SNPs) were identified during the early tillering stage, while five (21 SNPs) were identified during the active tillering stage (Table I; File S3).

Interestingly, nearly a half (3 QTL) of the QTL persisted across developmental stages, suggesting that they may impact growth throughout the vegetative phase captured by the two experiments. For instance, the most significant persistent association was identified in a region on chromosome 6 (~26.7.2-26.6 Mb) (Fig. 3). Significant signals were identified for 16 of the 38 time points, with the earliest association observed at 22 and the latest at 39 DAT. This QTL explained only a small portion of the percent of the phenotypic variation for PSA at 35 DAT (~5%). Several QTL were identified that displayed a time specific effect. A single QTL at ~5.9 Mb on chromosome 4 was associated with PSA during the early tillering stage at 14 and from 17-26 DAT. Two QTL had effects on PSA only during the active tillering stage. Of these two, a QTL located at ~3 Mb on chromosome 8 had the largest effect on PSA, and explained approximately 10% of the variation for PSA at 19 DAT (this QTL was only identified during the active tillering stage). This QTL was associated with PSA at 19-22 DAT. These results indicate that shoot biomass may be regulated by multiple loci with small effects that act in both a transient and persistent manner throughout early vegetative growth.

4.2.3 Functional GWAS analysis

While GWAS of PSA at discrete intervals provides information regarding the genetic basis of plant size over time, modeling plant growth allows for the data to be reduced to a mathematical equation that is defined by a small number of parameters, which can then be used as traits for

genetic analysis. This approach allows for the detection of genomic regions associated with growth trajectories, rather than discrete estimates of plant size. To identify genetic regions associated with shoot growth dynamics, GWAS was performed using the model parameters that were obtained from modeling shoot biomass accumulation at the early tillering and active tillering stages separately. A total of 19 SNPs ($p < 2.33 \times 10^{-5}$), corresponding to 7 unique QTL were identified for model parameters at both developmental stages (Table II; Fig. 4). The early tillering stage displayed a higher number of associations, with 5 QTL detected, while two QTL were detected for the model parameter M_o during the active tillering stage. For all significant markers the percent variation explained by individual QTL was low (≤ 0.10). The most significant association was detected at ~ 25.2 Mb on chromosome 2 for r during the early tillering stage and explained approximately 6% of total variation for this trait. No QTL were identified that were common across developmental stages, suggesting that shoot growth dynamics during the early and active tillering stages may be regulated by distinct genetic mechanisms.

In the power-law function, the parameters r and β are used to control the rate of growth over time. Both model parameters displayed strong phenotypic correlation during both the early and active tillering stages (-0.93, -0.54, respectively). Joint analysis of correlated phenotypes may improve power to detect genetic associations in GWAS compared to univariate approaches (Ferreira and Purcell, 2009; Kim and Xing, 2009; Korte et al., 2012; O'Reilly et al., 2012; Zhou and Stephens, 2014). With this in mind, model parameters r and β, which control growth rate, were used in a bivariate mixed model to detect associations in the early and active tillering stages (Zhou and Stephens, 2014). Overall the bivariate approach identified nine QTL (19 SNP) in total compared to the univariate approach, however more than twice as many QTL were identified

during the active tillering stage using the bivariate approach (Fig. 4 D, H). During the early tillering stage four QTL were identified, two of which were identified using only the bivariate approach. While at the active tillering stage five QTL were identified, all of which were detected with the bivariate approach. The most significant association in the early tillering stage was detected at ~10.1 Mb on chromosome 8 (6.11×10^{-8}; wd8001592). These results suggest that the bivariate approach is more effective for detecting loci with small effects that regulate shoot growth dynamics.

4.2.4 Shoot growth dynamics during early tillering may be regulated by gibberellic acid

To identify candidate genes that may regulate vegetative growth dynamics in rice, we treated significant SNPs within a 200kb window of one another as a single QTL, and genes within 200kb of each QTL were used for further analysis. The 200kb window was selected based on the estimated LD decay in this diversity panel (Zhao et al., 2011; Famoso et al., 2011). A gene encoding a GA2-oxidase protein (*OsGA2ox7*; LOC_Os02g41954), which is involved in GA catabolism, was identified within the QTL at ~25.2 Mb on chromosome 2 that was associated with *r* during the early tillering stage. *OsGA2ox7* is located approximately 54 kb away from the most significant SNP in this region, which is within the estimated LD within this region (60 kb). To examine the differences between allelic groups at this QTL, we performed RNA sequencing and GA quantification on a subset of accessions (File S5; Table S1). A linear model was fit in the Bioconductor package limma to examine the differences in expression between the two allelic groups at the most significant SNP for this QTL (Gentleman et al., 2004; Ritchie et al., 2015). Interestingly, the accessions in the major allelic group (A) displayed faster growth rate and significantly lower expression of this transcript ($p = 0.005$, SNP id2010960) compared to minor allelic groups (G) (Fig 5 A-C; File S6).

To determine whether there were differences in GA content between allelic groups at the QTL harboring *OsGA2ox7*, we quantified GA1, GA4 and GA20 levels in shoot tissue of 12 accessions within the major and minor allelic groups at this QTL. Samples were collected during the early tillering stage (10 DAT) to replicate the developmental timing of this QTL. Significant differences were observed between allelic groups for GA4, with the faster growing A allele displaying higher levels of GA4 ($p < 0.039$; Fig. 6A). Although slightly higher levels of GA1 were also observed in the A allelic groups, the differences were not significant at the chosen α level (α=0.05; Fig. 6B). These results indicate that natural variation for shoot growth dynamics during the early tillering stage may be partially explained by differences in bioactive GA levels. The differences in GA levels may be due to higher expression of the GA catalytic enzyme, *OsGA2ox7*.

4.2.5 Prediction of shoot growth dynamics and PSA

Since the QTL identified using GWAS explained only a small portion of the total variance for PSA and model parameters, an approach that captures the small effects of many markers, such as genomic selection, may be more advantageous for trait improvement than single marker strategies. With this in mind, genomic selection (GS) analysis was performed using RR-BLUP (ridge regression best linear unbiased predictor) to examine the potential for improving shoot growth dynamics and PSA in rice (Endelman, 2011). The accuracies of genomic selection were assessed using a five-fold cross validation using 36,901 SNPs, as well as 14 sets of varying size of the top SNPs identified from GWAS with 20 iterations for each SNP set. For each accession in the training population, the GEBVs were calculated using marker effects estimated in the training population and was correlated with phenotypes for the 72 accessions in the validation

population. The relationship between GEBV calculated with all 36,901 markers and observed phenotypes is presented in Fig 7 and Figs S1-S28.

Prediction accuracies ranged from 0.39 to 0.73 (averaged across all SNP sets) with the highest accuracy observed for PSA at 30 DAT (0.73; averaged across all SNP sets; File S7). The mean and standard deviation across the twenty iterations is provided as Supplemental File S7. Of the three model parameters, M_o exhibited the highest accuracies (0.55 and 0.61) during the early and active tillering stages, respectively (Fig. 8; File S7). Although traits could be predicted with reasonable accuracy using the top 50 most significant SNPs from GWAS (0.37-0.68), the accuracy of prediction improved by including more markers. These trends in accuracy are consistent with a polygenic architecture where hundreds to thousands of loci contribute small effects to the phenotype (Kooke et al., 2016). These results indicate that GS may be an effective strategy for improving shoot biomass and enhancing shoot growth dynamics during the vegetative growth stage in rice.

4.3 Discussion

Growth is a complex phenotype that is greatly influenced by environmental and developmental cues. The observable phenotype is the cumulative result of many biological processes that occur over time. In grasses, the timing of developmental switches that initiate tiller development or transition to the reproductive phase has large influence on vigor and plant size. The temporal nature of such genetic effects are evidenced by the transient QTL associated with PSA, as well as the small overlap in QTL associated with model parameters across developmental stages. With the advent of high-throughput phenomics platforms, high-resolution temporal data can be collected non-destructively for large mapping populations. Mathematical equations that describe a developmental or physiological process can be applied to this data to reduce temporal growth

trajectories to a few mathematical parameters, and may capture components of the phenotype that may not be detected through cross-sectional phenotyping approaches. The additional phenotypic information provided using this approach is supported by the temporal correlation trend observed between model parameters and PSA, as well as the identification of 11 unique QTL that were identified only with the functional GWAS approach. Only four QTL were identified with both the cross sectional and functional GWAS approaches. The identification of 11 QTL that are unique to model parameters may be due to the greater power to detect alleles with small effects by utilizing phenotypic data across time points, or the model parameters may be elucidating component(s) of growth phenotype not intuitively derived from discrete measurements of plant size.

The influence of plant hormones on agronomic, growth-related traits is well document in rice and other cereals (Peng et al., 1999; Yamamuro et al., 2000; Ikeda et al., 2001; Chandler et al., 2002; Chono et al., 2003; Li et al., 2003; Nakamura et al., 2006; Sakamoto, 2006). Gibberellic acid is a key regulator of plant growth and development, and has contributed substantially to grain production in the 20th century through the development of high-yielding, semi-dwarf cereals (Peng et al., 1999; Sasaki et al., 2002). In rice, the most widely used gene for modifying plant architecture in modern varieties is *sd1*, which encodes a 20-oxidase GA biosynthetic protein (Sasaki et al., 2002). A single nonsynonomous substitution within this gene has a drastic effect on plant architecture, resulting in a deficiency of bioactive GA and dwarf stature (Sasaki et al., 2002). In our study, GWAS provided evidence for GA in the regulation of early vigor, however the effects of the QTL were minor compared to *sd1*. A QTL at ~25.2 Mb on chromosome 2 was associated with *r* during the early tillering stage, and harbored a GA catabolic gene, *OsGA2ox7* (Busov et al., 2003; Lo et al., 2008; Rieu et al., 2008; Wuddineh et

al., 2015). Lo *et al.* showed that higher expression of *OsGA2ox7,* relative to wild-type plants, resulted in reduced stature and growth rate (Lo et al., 2008). In our study, accessions belonging to the major allele group displayed significantly higher expression of this gene, slower growth rate during the early tillering stage and lower GA4 levels, which is consistent with the previous report for *OsGA2ox7* (Lo et al., 2008).

Improving early vigor has remained a major challenge in rice breeding programs due to the complex genetic basis and large genotype-by-environment effect. GWAS has proven to be an indispensible approach to identify causal genes underlying traits in a variety of different organisms. GWAS are particularly effective for traits that are regulated by a few loci with large effects (Korte and Farlow, 2013). Several loci were identified for model parameters and PSA that harbored genes known to regulate growth in rice and other species, but despite the moderate to high heritability observed for the traits, each individual locus only explained a small portion of the phenotypic variation for the trait. These patterns are typical of complex polygenic traits. With this type of genetic architecture, single marker strategies (i.e. marker-assisted selection) may not be the most effective approach for genetic improvement. Genomic selection, on the other hand, does not focus on individual markers, and thus is a more effective approach to breed for traits controlled by many loci with small effects compared to marker-assisted selection (Jannink et al., 2010). Prediction accuracies for PSA and model parameters were relatively high, ranging from 0.39 to 0.73, which indicates that GS may be an effective strategy for improving shoot growth dynamics in rice. Moreover, high prediction accuracies were obtained using a subset of informative markers from association mapping, which may further reduce genotyping costs.

This study provides insights into the complex genetic architecture and molecular mechanisms underlying early shoot growth dynamics in rice. Although, early vigor is of interest in many

breeding programs, the complex genetic basis, temporal component of the phenotype and large genotype-by-environment effects have hindered genetic improvement. The evaluation of large mapping populations in a controlled environment using non-destructive phenomics provides high-resolution phenotypic data and reduces the influence of environment on this complex trait. Moreover the use of a large diversity panel allows for a greater allelic diversity to be queried, thus providing broader insight into the genetic basis of this trait compared to studies using biparental populations. Our study presents a foundational approach for the elucidation of vegetative growth dynamics and early vigor in rice. The presented approach of combining high-resolution image-based phenotyping coupled to functional mapping and genome prediction can be widely applicable for complex traits across numerous crop species.

4.4 Materials and Methods

4.4.1 Plant Materials and genotyping

A rice diversity panel (RDP1) consisting of 413 accessions was obtained from the USDA-ARS Dale Bumpers Rice Research Center and purified through single seed descent before they were phenotyped (Zhao et al., 2011; Famoso et al., 2011; Eizenga et al., 2014). After removing accessions with low seed or poor seed quality, a subset of 360 and 361 accessions were phenotyped during the early and late tillering stages respectively. All accessions of RDP1 were genotyped using 44k SNP array (Zhao et al., 2011; https://ricediversity.org/data/sets/44kgwas/). The genotypic dataset consisted of 33,901 markers for 360 accessions. Missing markers were imputed using Beagle with 20 iterations (Browning and Browning, 2016).

4.4.2 Greenhouse growth conditions and phenotyping

Two experiments were conducted at the Plant Accelerator, Australian Plant Phenomics Facility, at the University of Adelaide, AU. The first experiment consisted of 361 accessions and was repeated three times from Aug-Nov 2013. The experiment consisted of two smarthouses that were used consecutively for three periods, with each period forming a block. In each smarthouse, 216 pots were positioned across 24 lanes. A partially replicated design was used for each period. The plants were phenotyped from 8-27 DAT. A complete description of the experimental design is provided in Campbell *et al.* (Campbell et al., 2015). The second experiment was replicated three times from Sept 2014-Dec 2014. The greenhouse conditions and experimental design was nearly identical to that described in above. However, only 360 accessions were used due to seed availability. The plants were phenotyped from 19-41 DAT. For each experiment the plants were imaged daily using VIS/RGB camera (Basler Pilot piA2400-12gc) from two side view angles separated by 90° and a single top view. A total of 142,671 images were generated for the 2013 experiment, while 152,997 images were generated from the 2014 experiment.

4.4.3 Image Processing

All 295,668 images were processed using ImageHarvest (IH) (Knecht et al., 2016). Two processing pipelines were developed to extract plant pixels from side-view and top-view images. Briefly, the side-view processing pipeline consists of three major steps: (1) preprocessing, (2) thresholding, and (3) removal of the pot and carrier from the image. Preprocessing smooths the image providing a more uniform background and was achieved using the `gaussianBlur` function in IH. The image was converted to greyscale and the `adaptiveThreshold` function was used to remove the majority of the non-plant pixels from the image. `AdaptiveThreshold` separates the image into smaller windows and, in each window, a

threshold is applied and pixels are removed based on their intensity. A region of interest was defined around the pot and mean shift segmentation was applied using the `meanshift` function in IH to extract the plant pixels from this region.

Top view images were processed using a slightly different pipeline. First, to create a more uniform coloring in the image, the color values were normalized using the `normalizeByIntensity` function in IH. For each pixel the `normalizeByIntensity` function rescales the color value for each channel (R, G, B) based on its intensity, which is defined as the sum of all color channels (R + G + B) for that pixel. Next, the image was segmented using the `meanshift` function. Plant pixels were extracted from the image if the value for the G channel was greater than the value for the R channel. Finally, a morphological closing operation consisting of a dilation and erosion step was used to fill in "holes" in the plant that may have been caused by the processing steps. This was achieved using the morphology function in IH.

4.4.4 Statistical Analysis of projected shoot area (PSA)

To quantify shoot growth at each time point, the plant pixels extracted from each of the three images (side view 1, side view 2 and top view) were summed. This metric, here defined as projected shoot area (PSA) has been shown to be a reliable estimate of shoot biomass (Berger et al., 2010; Campbell et al., 2015). For each period (i.e. Aug-Nov 2013 and Sept-Dec 2014), PSA was combined across experiments, and a linear model was fitted to calculate the adjusted means for each accession using the 'lsmeans' function in the LSMeans package in R (R Core Team, 2014; https://cran.r-project.org/web/packages/lsmeans/index.html). The experiment and treatment were considered fixed effects, and accession as a random effect in the linear model. The adjusted means were used for cross-sectional GWAS of PSA and functional modeling.

4.4.5 Functional modeling of temporal trends in PSA

For each accession the adjusted mean for PSA was modeled using the following power-function (Paine et al., 2012):

$$M_t = \left(M_0{}^{(1-\beta)} + rt(1-\beta)\right)^{1/(1-\beta)}$$

where M_t is the PSA at time t, M_o is the PSA at the start of imaging, and r and β are parameters controlling the growth rate. The exponent β allows the relative growth rate to slow as biomass increases (Paine et al., 2012). The choice of model parameters is not trivial in non-linear regression, and is often based on visual inspection or *a priori* knowledge regarding the nature of the data. Therefore to reduce the labor burden of fitting hundreds of models, a three-step approach was used to obtain optimal estimates of the parameters β, r, and M_o using the nls2, nlmrt and nlme packages in R (Grothendieck, 2013; Nash, 2013; Pinheiro et al., 2015). First, rough starting estimates were obtained using nls2 with the brute-force algorithm with 10,000 iterations. Nls2 performs a grid search to obtain rough estimates of the model parameters, using a range of parameter values supplied by the user. The best estimates for each of the model parameters were supplied as starting values to the nlxb function in nlmrt. Nlmrt can be more robust in finding solutions compared to nls, especially when the data to be modeled has small or zero residuals. Finally, the results obtained from nlxb were used to fit the model in nls.

4.4.6 Genome-wide association analysis of PSA

To identify genomic regions associated with projected shoot area (PSA) at each time point, a mixed model that accounted for kinship and population structure was used for GWAS using the EMMA algorithm (Kang et al., 2008). The mixed linear model can be summarized as $y = X\beta + C\gamma + Zu + e$ where y is a vector of phenotype; β is a vector of fixed marker effects; γ is a vector

of principal component (PC) effects fit in order to account for population structure; u is a vector of polygenic effects caused by relatedness; e is a vector of residuals; X is a marker incidence matrix relating β to y; C is an incidence matrix relating γ to y which consists of the first four principal components (PCs) resulting from a PC analysis; Z is the corresponding design matrix relating y to u. It is assumed $\mathbf{u} \sim MVN(0, \mathbf{K}\sigma_u^2)$ and $\mathbf{e} \sim MVN(0, \mathbf{I}\sigma_e^2)$ where K is a kinship matrix estimated using an allele-sharing matrix calculated from the SNP data. Markers with a minor allele frequency less than 5% were excluded from the analysis. The parameter β during the active tillering stage was log transformed to provide a normal distribution prior to association mapping.

To test for genetic associations with the model parameters M_o, β and r in the early and active tillrering stages both univaraite and multivariate approaches were implemented using Genome-wide Efficient Mixed Model Association (GEMMA) (Zhou and Stephens, 2014). To identify SNPs associated with the two power-law parameters describing growth rate, β and r, a bivariate mixed model was used. The mixed model that was implemented is similar to that described above. Population and relatedness were accounted for using the top four PCs and a centered genetic relatedness matrix. For each SNP, a likelihood ratio test was used to test the alternative hypothesis H1: $\beta \neq 0$, against the null hypothesis H0: $\beta = 0$.

To account for multiple testing, the Sidak correction using the effective number of tests (M_{eff}) was applied (Li and Ji, 2005). Briefly, the effective number of independent tests (M_{eff}) was determined by eigenvalue decomposition of the correlation matrix among 34,960 SNPs (MAF < 0.05 for 360 accessions). The test criteria was then adjusted using the M_{eff} with the Sidak (1967) correction below

$\alpha_p = 1 - (1- \alpha_e)^{1/Meff}$,

where α_p is the comparison-wise error rate and α_e is the experiment-wise error rate ($\alpha = 0.05$). Using this approach Meff was determined to be 2,203, which is the number of eigenvalues necessary to explain 99% of variation, and α_p: 2.33×10^{-5}.

SNPs within 200 Kb, which represents the estimated LD decay in this population were considered as a single QTL (Zhao et al., 2011; Famoso et al., 2011). All genes within 200 Kb of significant SNPs were considered as potential candidate genes. Analysis of variance (ANOVA) was used to estimate proportion of phenotypic variance accounted for by significant SNPs after adjusting for population structure effects. ANOVA was used to compare linear models $y= X\beta + C\gamma + e$ and $y= C\gamma + e$ where y is a vector of phenotype; β is the SNP effect; γ is a vector of principal component (PC) effects; e is a vector of residuals; X is the SNP genotype; C is a matrix of the first four principal components (PCs).

4.4.7 Genomic Selection

Genomic selection was performed using RR-BLUP implemented in the rrBLUP package in R (Endelman, 2011). Prediction was performed using all 36,901 SNPs as well as various subsets of the most significant associations from GWAS (GWAS-informed prediction). The inclusion of SNPs was based on GWAS performed in the training population for each fold and replicate. The prediction accuracy was assessed using a five-fold cross validation. Accessions were randomly assigned to five folds, with four folds (288 accessions) used as a training population and the remaining 72 accessions used for validation. To assess prediction accuracy, genomic estimated breeding values (GEBVs) for the 72 accessions were calculated using the marker effects determined from the training population and were correlated to the observed phenotypes for those accessions. GWAS was performed using the `GWAS` function in the rrBLUP package with

the P3D option using the top four PCs were used to account for population structure and the realized additive relationship matrix, calculated with the `A.mat` function was used to account for cryptic relatedness between accessions (Zhang et al., 2010; Endelman, 2011). SNPs were ranked based on the $-\log_{10}(p)$ values and the top 50, 100, 200, 500, 1000, 2500, 5000, 7500, 10000, 15000, 20000, 25000 and 30000 markers were selected for prediction analysis. The predictive ability was calculated as the correlation between the GEBV of the validation population and the observed values for the validation population. Twenty iterations of the five-fold cross validation was performed for each trait.

4.4.8 Hormone Quantification

To quantify gibberellic acid levels, shoot tissue was collected from 12 accessions at 10 days after transplant (Table S1). This time point was selected to reflect the start of imaging during the early tillering stage experiment. The plants were cultured in a growth chamber with temperatures maintained at 28 °C and 25 °C in day and night respectively and 60% relative humidity. Lighting was maintained at 800 micromoles m^{-2} s^{-1} using high pressure sodium lights (Phillips). Seeds were geminated in 1/2 strength MS media (0.8% Agar) and transplanted to pots filled with Turface (Profile Products, LLC) four days after sowing. Four days after transplant, a 1/2 strength Yoshida solution was provided and pH was maintained at 5.8 (Yoshida et al., 1976). Two plants were grown in each pot. Shoot tissue was from two plants was pooled, flash frozen and ground in liquid nitrogen. Hormones were extracted from 250 mg tissue (fresh weight) using published extraction protocol and was quantified using high-performance liquid chromatography–mass spectrometry (Pan et al., 2010). GA1 and GA4 were used as internal standards.

4.4.9 Growth conditions for transcriptome experiment

For gene expression analysis plants from 31 accessions were grown in a controlled environment growth chamber (Table S1). Growth conditions are identical to those described above. At 10 DAT, aerial parts of the seedlings were excised from the roots and frozen immediately in liquid nitrogen. The samples were ground with Tissuelyser II (Invitrogen) and total RNA was isolated with RNAeasy isolation kit (Qiagen) according to manufacturer's instructions. On-column DNAse treatment was performed to remove genomic DNA contamination (Qiagen). Sequencing was performed using Illumina HiSeq 2500. Sixteen RNA samples were combined in each lane. Two biological replicates were used for each accession (Table S1).

4.4.10 RNA-seq mapping and analysis

After being examined using the package FastQC, Short reads, obtained from Illumina 101-bp length single-end RNA sequencing, were screened and trimmed using Trimmomatic to ensure each read has average quality score larger than 30 and longer than 15bp (Andrews, 2010; Bolger et al., 2014). The trimmed short reads were mapped against to the rice genome (*Oryza sativa* 'Nipponbare' MSU Release 6.0) using TopHat (v.2.0.10), allowing up to two base mismatches per read. Reads mapped to multiple locations were discarded (Trapnell et al., 2009). The number of reads for each gene in the MSU v6 annotation was counted using HTSeq and the "union" resolution mode was used (Anders et al., 2014).

The Bioconductor packages 'limma' and 'EdgeR' were used to identify genes that were differentially expressed between allelic groups at id3013397 and id2010960 (Gentleman et al., 2004; Robinson and Oshlack, 2010; Ritchie et al., 2015). Briefly, genes with a total read count below 50 across all 64 samples were excluded from the differential expression analysis. The libraries were normalized using the trimmed mean of M-values (TMM) in 'EdgeR' (Robinson

and Oshlack, 2010). The voom approach, implemented in ‘limma’, was used to identify differentially expressed genes between the two allelic groups for the SNPs id3013397 and id2010960 (Law et al., 2014). Benjamini and Hochberg’s method was used to control the false discovery rate, and genes with an adjusted p-value ≤ 0.1 were considered differentially expressed (Benjamini and Hochberg, 1995).

4.5 References

Anders, S., P.T. Pyl, and W. Huber. 2014. HTSeq – A Python framework to work with high-throughput sequencing data. Bioinformatics 31(2): 166–169.

Andrews, S., and others. 2010. FastQC: A quality control tool for high throughput sequence data.

Bac-Molenaar, J. a., D. Vreugdenhil, C. Granier, and J.J.B. Keurentjes. 2015. Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. J. Exp. Bot.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B: 289–300.

Berger, B., B. Parent, and M. Tester. 2010. High-throughput shoot imaging to study drought responses. J. Exp. Bot. 61(13): 3519–28.

Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics: btu170.

Browning, B.L., and S.R. Browning. 2016. Genotype imputation with millions of reference samples. Am. J. Hum. Genet. 98(1): 116–126.

Busemeyer, L., A. Ruckelshausen, K. Möller, A.E. Melchinger, K. V Alheit, H.P. Maurer, V. Hahn, E. a Weissmann, J.C. Reif, and T. Würschum. 2013. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. Sci. Rep. 3: 2442.

Busov, V.B., R. Meilan, D.W. Pearce, C. Ma, S.B. Rood, and S.H. Strauss. 2003. Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature. Plant Physiol. 132(3): 1283–1291.

Cairns, J.E., O.S. Namuco, R.Torres, F. a. Simborio, B. Courtois, G. a. Aquino, and D.E. Johnson. 2009. Investigating early vigour in upland rice (Oryza sativa L.): Part II. Identification of QTLs controlling early vigour under greenhouse and field conditions. F. Crop. Res. 113(3): 207–217.

Campbell, M.T., A.C. Knecht, B. Berger, C.J. Brien, D. Wang, and H. Walia. 2015. Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. Plant Physiol. 168(August): pp.00450.2015.

Chandler, P.M., A. Marion-Poll, M. Ellis, and F. Gubler. 2002. Mutants at the Slender1 locus of barley cv Himalaya. Molecular and physiological characterization. Plant Physiol. 129(1): 181–190.

Chen, D., K. Neumann, and S. Friedel. 2014. Dissecting the Phenotypic Components of Crop Plant Growth and Drought Responses Based on High-Throughput Image Analysis. Plant Cell … 26(12): 4636–4655.

Chono, M., I. Honda, H. Zeniya, K. Yoneyama, D. Saisho, K. Takeda, S. Takatsuto, T. Hoshino, and Y. Watanabe. 2003. A semidwarf phenotype of barley uzu results from a nucleotide

substitution in the gene encoding a putative brassinosteroid receptor. Plant Physiol. 133(3): 1209–1219.

Dai, M., Y. Hu, Q. Ma, Y. Zhao, and D.-X. Zhou. 2008. Functional analysis of rice HOMEOBOX4 (Oshox4) gene reveals a negative function in gibberellin responses. Plant Mol. Biol. 66(3): 289–301.

Eizenga, G.C., M. Ali, R.J. Bryant, K.M. Yeater, A.M. McClung, S.R. McCouch, and others. 2014. Registration of the rice diversity panel 1 for genomewide association studies. J. Plant Regist. 8(1): 109–116.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4(3): 250–255.

Famoso, A.N., K. Zhao, R.T. Clark, C.-W. Tung, M.H. Wright, C. Bustamante, L. V Kochian, and S.R. McCouch. 2011. Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7(8): e1002221.

Ferreira, M.A.R., and S.M. Purcell. 2009. A multivariate test of association. Bioinformatics 25(1): 132–133.

Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, and others. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5(10): 1.

Golzarian, M.R., R. a Frick, K. Rajendran, B. Berger, S. Roy, M. Tester, and D.S. Lun. 2011. Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods 7(1): 2.

Grothendieck, G. 2013. nls2: Non-linear regression with brute force. R Packag. version 2.

Hairmansis, A., B. Berger, M. Tester, and S.J. Roy. 2014. Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice. Rice 7(1): 16.

Honsdorf, N., T. March, B. Berger, M. Tester, and K. Pillen. 2014. High-Throughput Phenotypig to Detect Drought Tolerance QTL in Barley Introgression Lines. PLoS One 9(5): e97047.

Ikeda, A., M. Ueguchi-Tanaka, Y. Sonoda, H. Kitano, M. Koshioka, Y. Futsuhara, M. Matsuoka, and J. Yamaguchi. 2001. slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. Plant Cell 13(5): 999–1010.

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9(2): 166–177.

Kang, H.M., N. a Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178(3): 1709–23.

Kim, S., and E.P. Xing. 2009. Statistical estimation of correlated genome associations to a

quantitative trait network. PLoS Genet 5(8): e1000587.

Knecht, A.C., M.T. Campbell, A. Caprez, D.R. Swanson, and H. Walia. 2016. Image Harvest: an open-source platform for high-throughput plant image processing and analysis. J. Exp. Bot. 67(11): 3587–3599Available at http://jxb.oxfordjournals.org/cgi/content/long/erw176v1.

Kooke, R., W. Kruijer, R. Bours, F.F.M. Becker, A. Kuhn, J. Buntjer, T. Doeswijk, J. Guerra, H.J. Bouwmeester, D. Vreugdenhil, and others. 2016. Genome-wide association mapping and genomic prediction elucidate the genetic architecture of morphological traits in Arabidopsis thaliana. Plant Physiol.: pp--00997.

Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9(1): 1.

Korte, A., B.J. Vilhjálmsson, V. Segura, A. Platt, Q. Long, and M. Nordborg. 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat. Genet. 44(9): 1066–1071.

Law, C.W., Y. Chen, W. Shi, and G.K. Smyth. 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15(2): 1.

Li, J., and L. Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). 95(3): 221–227.

Li, X., Q. Qian, Z. Fu, Y. Wang, G. Xiong, D. Zeng, X. Wang, X. Liu, S. Teng, F. Hiroshi, and others. 2003. Control of tillering in rice. Nature 422(6932): 618–621.

Liu, L., Y. Lai, J. Cheng, L. Wang, W. Du, Z. Wang, and H. Zhang. 2014. Dynamic Quantitative Trait Locus Analysis of Seed Vigor at Three Maturity Stages in Rice. PLoS One 9(12): e115732.

Lo, S.-F., S.-Y. Yang, K.-T. Chen, Y.-I. Hsing, J.A.D. Zeevaart, L.-J. Chen, and S.-M. Yu. 2008. A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. Plant Cell 20(10): 2603–2618.

Lu, X.-L., A.-L. Niu, H.-Y. Cai, Y. Zhao, J.-W. Liu, Y.-G. Zhu, and Z.-H. Zhang. 2007. Genetic dissection of seedling and early vigor in a recombinant inbred line population of rice. Plant Sci. 172(2): 212–220.

Luquet, D., J.C. Soulié, M.C. Rebolledo, L. Rouan, a. Clément-Vidal, and M. Dingkuhn. 2012. Developmental Dynamics and Early Growth Vigour in Rice 2. Modelling Genetic Diversity Using Ecomeristem. J. Agron. Crop Sci. 198(5): 385–398.

Mahender, A., A. Anandan, and S.K. Pradhan. 2015. Early seedling vigour, an imperative trait for direct-seeded rice: an overview on physio-morphological parameters and molecular markers. Planta 241(5): 1027–1050.

Moore, C.R., L.S. Johnson, I.-Y. Kwak, M. Livny, K.W. Broman, and E.P. Spalding. 2013. High-throughput computer vision introduces the time axis to a quantitative trait map of a plant growth response. Genetics 195(3): 1077–86.

Nakagawa, H., A. Tanaka, T. Tanabata, M. Ohtake, S. Fujioka, H. Nakamura, H. Ichikawa, and M. Mori. 2012. Short grain1 decreases organ elongation and brassinosteroid response in rice. Plant Physiol. 158(3): 1208–1219.

Nakamura, A., S. Fujioka, H. Sunohara, N. Kamiya, Z. Hong, Y. Inukai, K. Miura, S. Takatsuto, S. Yoshida, M. Ueguchi-Tanaka, and others. 2006. The role of OsBRI1 and its homologous genes, OsBRL1 and OsBRL3, in rice. Plant Physiol. 140(2): 580–590.

Nash, J. 2013. Nash. nlmrt: Functions for nonlinear least squares solutions, 2013. R Packag. version 9(0): 50.

O'Reilly, P.F., C.J. Hoggart, Y. Pomyen, F.C.F. Calboli, P. Elliott, M.-R. Jarvelin, and L.J.M. Coin. 2012. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS One 7(5): e34861.

Paine, C.E.T., T.R. Marthews, D.R. Vogt, D. Purves, M. Rees, A. Hector, and L. a. Turnbull. 2012. How to fit nonlinear plant growth models and calculate growth rates: An update for ecologists. Methods Ecol. Evol. 3: 245–256.

Pan, X., R. Welti, and X. Wang. 2010. Quantitative analysis of major plant hormones in crude plant extracts by high-performance liquid chromatography--mass spectrometry. Nat. Protoc. 5(6): 986–992.

Peng, J., D.E. Richards, N.M. Hartley, G.P. Murphy, K.M. Devos, J.E. Flintham, J. Beales, L.J. Fish, A.J. Worland, F. Pelica, and others. 1999. "Green revolution"genes encode mutant gibberellin response modulators. Nature 400(6741): 256–261.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team. 2015. {nlme}: Linear and Nonlinear Mixed Effects Models.

R Core Team. 2014. R: A Language and Environment for Statistical Computing.

Rebolledo, M.-C., M. Dingkuhn, A. Clement-Vidal, L. Rouan, and D. Luquet. 2012a. Phenomics of rice early vigour and drought response: Are sugar related and morphogenetic traits relevant? Rice 5(1): 22.

Rebolledo, M.C., M. Dingkuhn, B. Courtois, Y. Gibon, a. Clément-Vidal, D.F. Cruz, J. Duitama, M. Lorieux, and D. Luquet. 2015. Phenotypic and genetic dissection of component traits for early vigour in rice using plant growth modelling, sugar content analyses and association mapping. J. Exp. Bot. 66(18): 5555–5566.

Rebolledo, M.C., M. Dingkuhn, P. Péré, K.L. McNally, and D. Luquet. 2012b. Developmental Dynamics and Early Growth Vigour in Rice. I. Relationship Between Development Rate (1/Phyllochron) and Growth. J. Agron. Crop Sci. 198(5): 374–384.

Redoña, E.D., and D.J. Mackill. 1996. Genetic variation for seedling vigor traits in rice. Crop Sci. 36(2): 285–290.

Rieu, I., S. Eriksson, S.J. Powers, F. Gong, J. Griffiths, L. Woolley, R. Benlloch, O. Nilsson, S.G. Thomas, P. Hedden, and others. 2008. Genetic analysis reveals that C19-GA 2-oxidation is a major gibberellin inactivation pathway in Arabidopsis. Plant Cell 20(9):

2420–2436.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.: gkv007.

Robinson, M.D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11(3): 1.

Sakamoto, T. 2006. Phytohormones and rice crop yield: strategies and opportunities for genetic improvement. Transgenic Res. 15(4): 399–404.

Sasaki, A., M. Ashikari, M. Ueguchi-Tanaka, H. Itoh, A. Nishimura, D. Swapan, K. Ishiyama, T. Saito, M. Kobayashi, G.S. Khush, and others. 2002. Green revolution: a mutant gibberellin-synthesis gene in rice. Nature 416(6882): 701–702.

Slovak, R., C. Göschl, X. Su, K. Shimotani, T. Shiina, and W. Busch. 2014. A scalable open-source pipeline for large-scale root phenotyping of Arabidopsis. Plant Cell Online 26(6): 2390–2403.

Sunohara, H., T. Kawai, S. Shimizu-Sato, Y. Sato, K. Sato, and H. Kitano. 2009. A dominant mutation of TWISTED DWARF 1 encoding an. ALPHA.-tubulin protein causes severe dwarfism and right helical growth in rice. Genes Genet. Syst. 84(3): 209–218.

Topp, C.N., A.S. Iyer-Pascuzzi, J.T. Anderson, C.-R. Lee, P.R. Zurek, O. Symonova, Y. Zheng, A. Bucksch, Y. Mileyko, T. Galkovskyi, B.T. Moore, J. Harer, H. Edelsbrunner, T. Mitchell-Olds, J.S. Weitz, and P.N. Benfey. 2013. 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. Proc. Natl. Acad. Sci. U. S. A. 110(18): E1695-704.

Trapnell, C., L. Pachter, and S.L. Salzberg. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9): 1105–11.

Wu, R., and M. Lin. 2006. Functional mapping—how to map and study the genetic architecture of dynamic complex traits. Nat. Rev. Genet. 7(March): 229–237.

Wuddineh, W.A., M. Mazarei, J. Zhang, C.R. Poovaiah, D.G.J. Mann, A. Ziebell, R.W. Sykes, M.F. Davis, M.K. Udvardi, and C.N. Stewart. 2015. Identification and overexpression of gibberellin 2-oxidase (GA2ox) in switchgrass (Panicum virgatum L.) for improved plant architecture and reduced biomass recalcitrance. Plant Biotechnol. J. 13(5): 636–647.

Würschum, T., W. Liu, L. Busemeyer, M.R. Tucker, J.C. Reif, E. a Weissmann, V. Hahn, A. Ruckelshausen, and H. Maurer. 2014. Mapping dynamic QTL for plant height in triticale. BMC Genet. 15: 59.

Xu, Z., X. Shen, and W. Pan. 2014. Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. PLoS One 9(8): e102312.

Yamamuro, C., Y. Ihara, X. Wu, T. Noguchi, S. Fujioka, S. Takatsuto, M. Ashikari, H. Kitano, and M. Matsuoka. 2000. Loss of function of a rice brassinosteroid insensitive1 homolog

prevents internode elongation and bending of the lamina joint. Plant Cell 12(9): 1591–1605.

Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang, W. Fang, H. Feng, W. Xie, X. Lian, G. Wang, Q. Luo, Q. Zhang, Q. Liu, and L. Xiong. 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat. Commun. 5: 5087.

Yoshida, S., D.A. Forno, and J.H. Cock. 1976. Laboratory Manual for Physiological Studies of Rice.

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and others. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42(4): 355–360.

Zhao, K., C.W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2: 467Available at http://www.ncbi.nlm.nih.gov/pubmed/21915109%5Cnhttp://www.nature.com/ncomms/journal/v2/n9/pdf/ncomms1467.pdf.

Zhou, X., and M. Stephens. 2014. Efficient multivariate linear mixed model algorithms for genome- wide association studies. Nat. Methods 11(4).

4.6 Figures

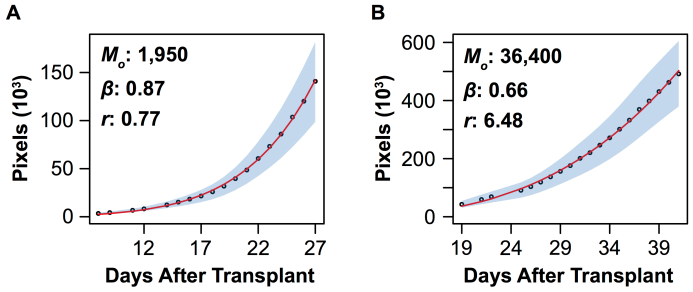


Fig. 1. Shoot growth trajectories of RDP1 during the early (A) and active tillering (B) stages. The mean growth across all RDP1 accessions was fit using a power-law function and is indicated by the red line. The standard deviation is indicated by the light blue shadow. The points indicate the mean growth across all accessions at each individual time point.

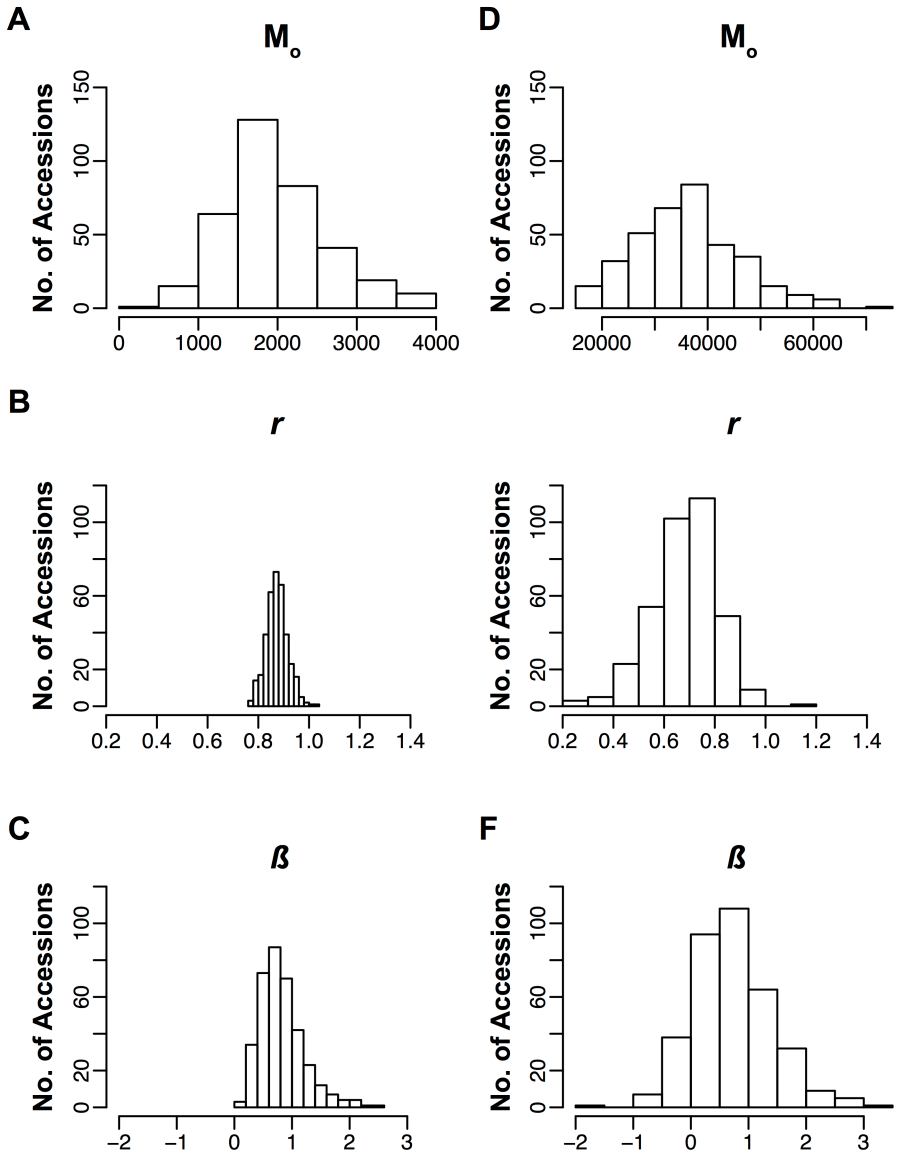


Fig. 2. Distribution for each model parameter during the early (A-C) and active (D-E) tillering stages. The parameters β and r control the rate of growth, while M_o represents the estimate of the starting biomass at the start of the experiment. Values for β during the active tillering stage (F) are log transformed.

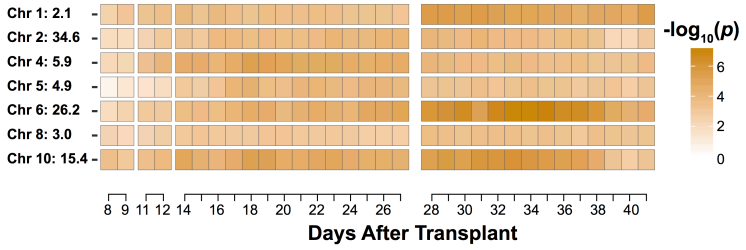


Fig. 3. Genome wide association analysis of PSA at each timepoint during the early and active tillering stages. GWAS was performed at each time point and significant SNPs ($p < 2.33 \times 10^{-5}$) within a 200 kb window were considered to be a single QTL. QTL positions are indicated on the left of the heatmap. QTLs detected during the overlapping time points are provided as Table S2.

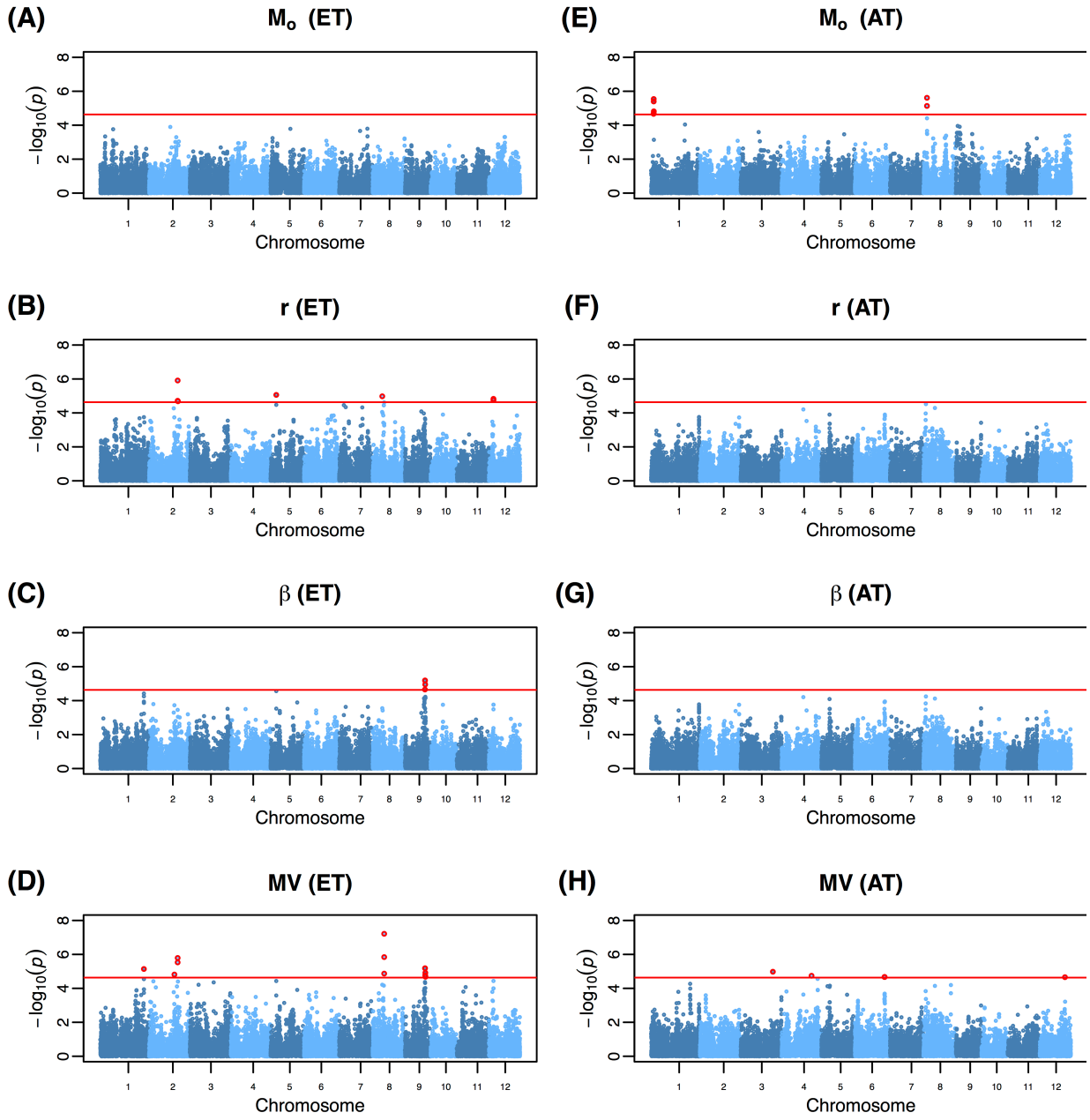


Fig. 4. Genome wide association analysis of the model parameters during the early (A-D) and active tillering (E-G) stages. The red horizontal line indicates the genome-wide significance level ($p < 2.33 \times 10^{-5}$) as determined using the Meff method with the Sidak

correction (Li and Ji, 2005). Significant SNPs are highlighted in red. MV: Multivariate; ET: Early tillering; AT: Active tillering

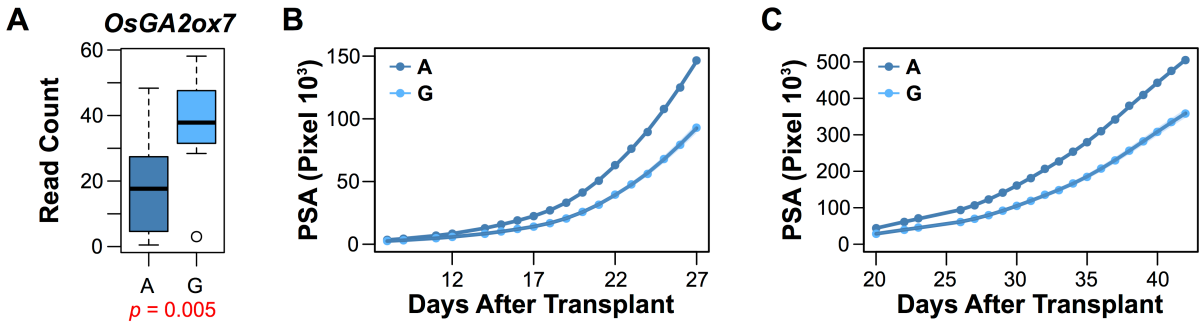


Fig. 5. Role of GA in the contrasting growth responses between allelic groups at ~25.2 Mb on chromosome 2. (A) Expression of *OsGA2ox7* between allelic groups at id2010960. (B, C) Mean growth trajectories for allelic groups at the most significant SNP associated with r (id2010960) during the early (B) and active (C) tillering stages. Statistical significance was determined using the Bioconductor packages limma and edgeR and the resulting p-value is indicated in red. Both genes displayed significance difference in expression between allelic groups at the corresponding SNP (FDR < 0.1).

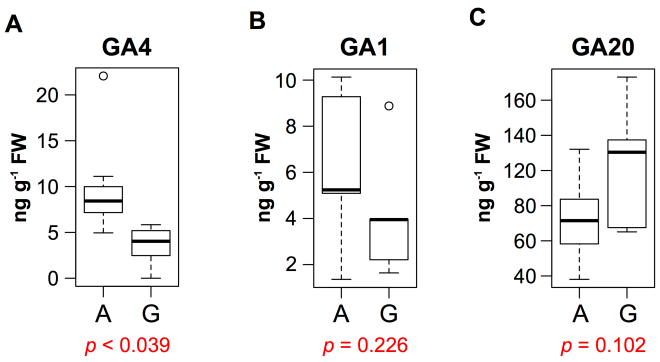


Fig. 6. Gibberellic acid content within allelic groups at id2010960. GA levels were quantified in shoot tissue at ~10 DAT. Differences between allelic groups was determined using significant was determined using Student's t-test and the resulting *p*-value is indicate in red (*n*=12).

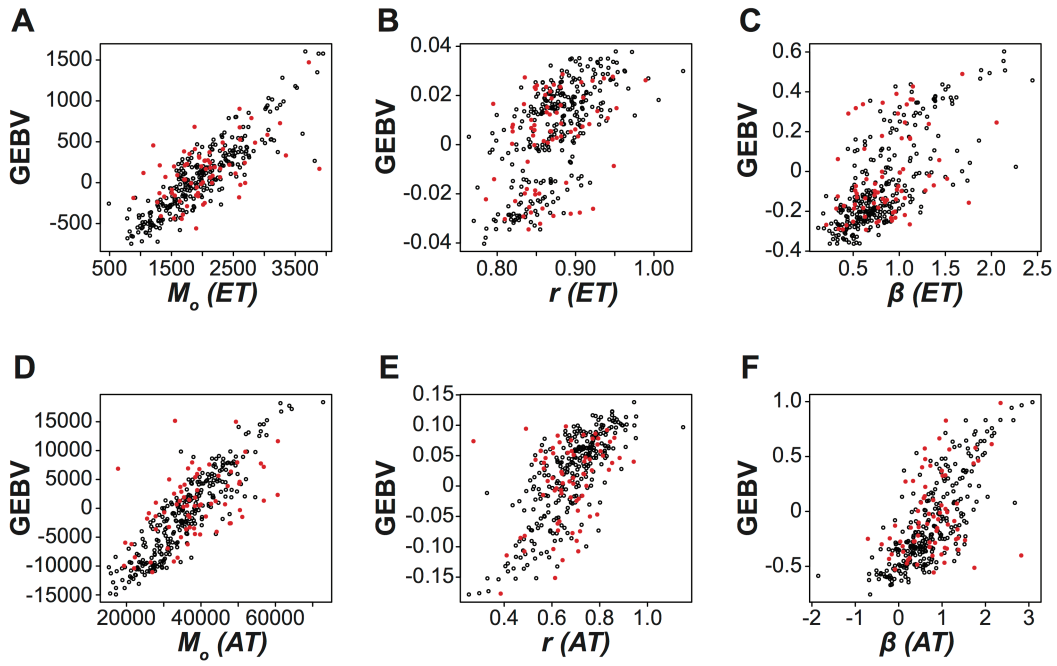


Fig. 7. Comparison of genomic estimated breeding values (GEBV) and observed phenotypes for model parameters at the early (A-C) and active tillering stages (D-F). GEBV were estimated using all 36,901 markers using a training population of 288 accessions. The black open points in each figure represent the accessions in the training population, while filled red points indicate accessions in the validation population (72 accessions). ET: early tillering; AT: active tillering

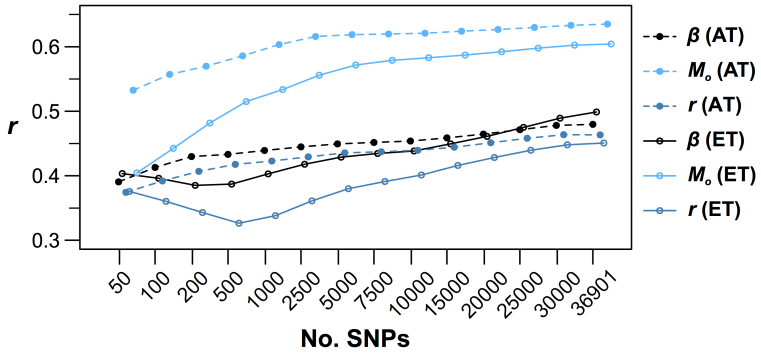


Fig. 8. Prediction accuracies for model parameters during the early and active tillering stages. Prediction accuracy was determined using a five-fold cross validation for each of the 14 sets of SNPs. The values represent the mean correlation between the GEBVs and the observed values for the validation population for 20 iterations of the five-fold cross validation.

4.7 Tables

Table I: QTL associated with projected shoot area (PSA) during the early and/or late vegetative stage. SNPs within 200 kb were merged and considered a single QTL. PSA: Projected shoot area

Chr.	QTL Position (bp)	Traits
1	2,136,478 – 2,334,214	PSA (Active Tillering)
2	34,616,145 – 34,630,711	PSA (Early Tillering); PSA (Active Tillering)
4	5,923,594 – 6,033,595	PSA (Early Tillering)
5	4,871,496 – 5,071,497	PSA (Early Tillering)
6	26,235,553 – 26,631,930	PSA (Early Tillering); PSA (Active Tillering)
8	2,901,247 – 3,101,248	PSA (Active Tillering)
10	15,397,674 – 15,659,792	PSA (Early Tillering); PSA (Active Tillering)

Table II: Subset of QTL associated with model parameters (M_o, r, β) during the early and/or late vegetative stage. SNPs within 200 kb were merged and considered a single QTL. ET: early tillering; AT: active tillering; MV: multivariate

Chr	QTL Position (BP)	Trait	Candidate Genes
1	2,136,478 – 2,334,214	M_o (AT)	
1	38,578,456 – 38,578,457	MV (ET)	*OsSD1*(Sasaki et al., 2002)
2	22,247,678 – 22,447,679	MV (ET)	
2	25,230,196 – 25,486,029	r (ET); MV (ET)	*OsGA2ox7* (Lo et al., 2008)
3	28,521,947 – 28,621,948	MV (AT)	
4	26,365,306 – 26,565,307	MV (AT)	
5	4,884,744 – 5,084,745	r (ET)	
6	26,277,256 – 26,366,993	MV (AT)	
8	2,901,212 – 3,101,248	M_o (AT)	
8	8,348,656 – 8,448,657	r (ET)	
8	10,138,560 – 10,145,645	MV (AT)	
9	17,857,737 – 18,361,253	β (ET); MV (ET)	*OsSG1* (Nakagawa et al., 2012), *OsHOX4* (Dai et al., 2008)
12	3,943,968 – 4,146,493	r (ET)	
12	22,231,779 – 22,331,780	MV (AT)	*OsTID1* (Sunohara et al., 2009)

4.8 Supplemental Materials

Fig S1.

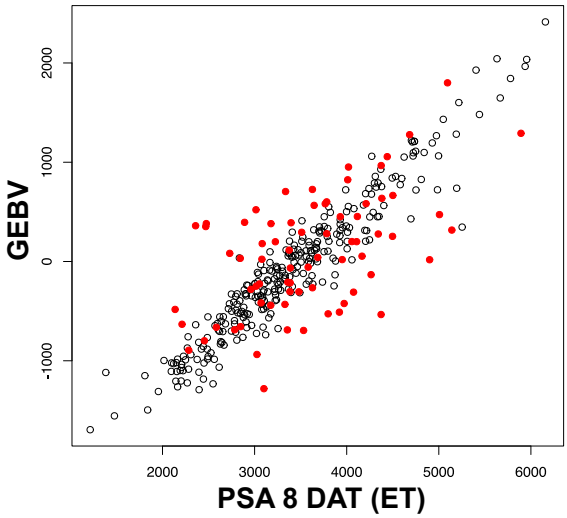


Fig S2.

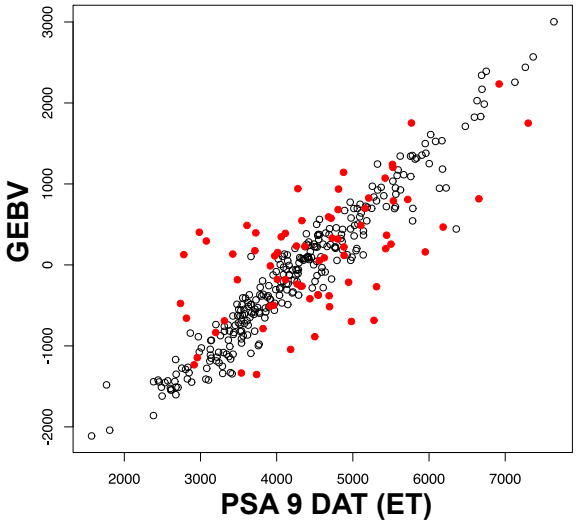


Fig S3.

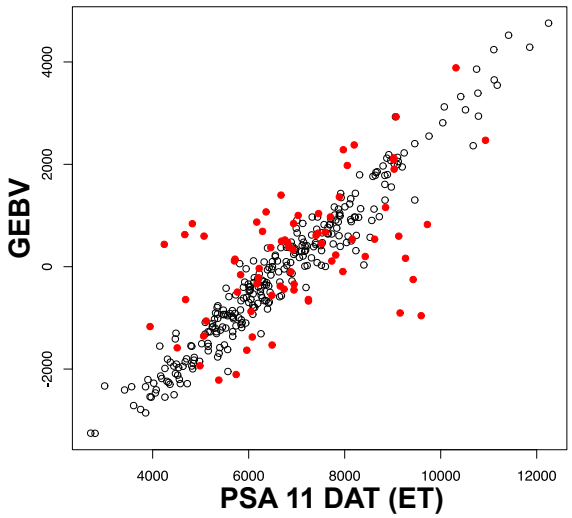

GEBV
4000
2000
0
-2000
4000
6000
8000
10000
12000
PSA 11 DAT (ET)

Fig S4.

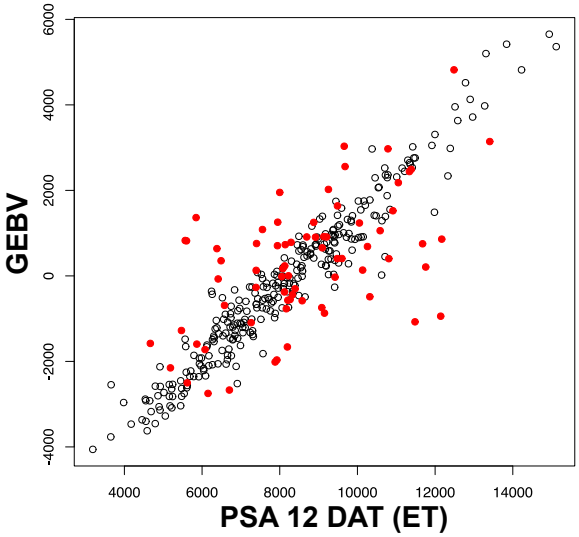

GEBV
6000
4000
2000
0
-2000
-4000
4000
6000
8000
10000
12000
14000
PSA 12 DAT (ET)

Fig S5.

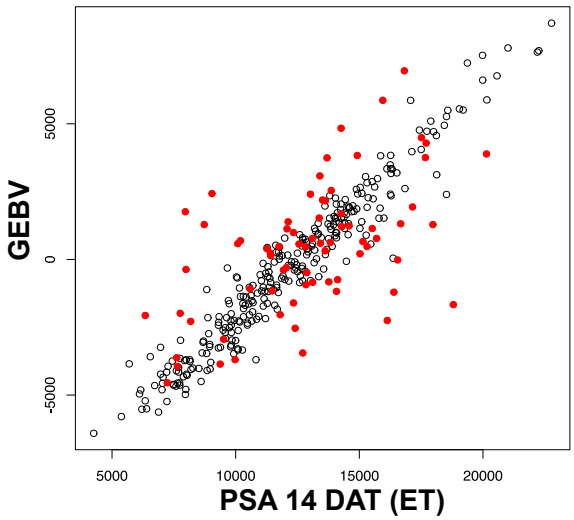

GEBV
5000
0
-5000
5000
10000
15000
20000
PSA 14 DAT (ET)

Fig S6.

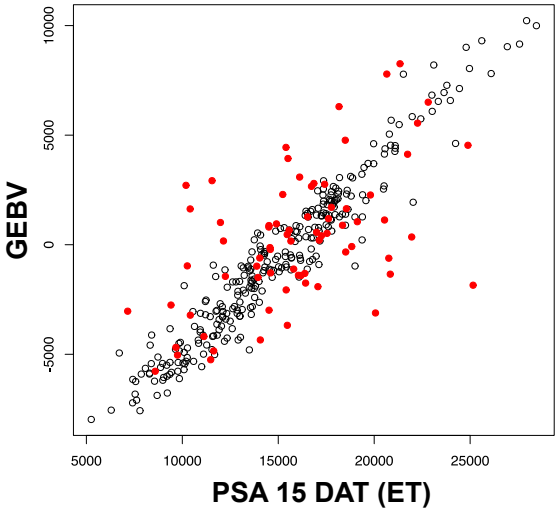

GEBV
10000
5000
0
-5000
5000
10000
15000
20000
25000
PSA 15 DAT (ET)

Fig S7.

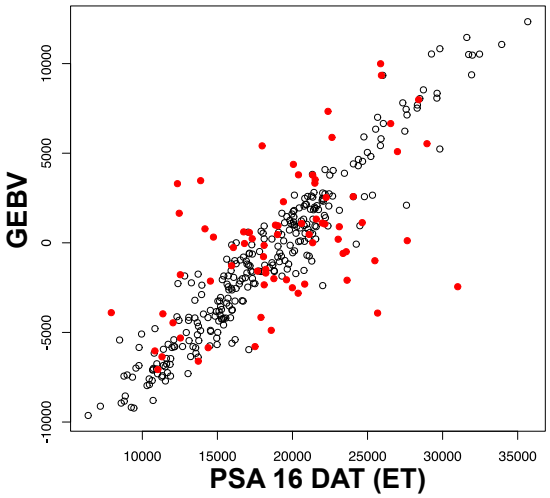

GEBV
10000
5000
0
-5000
-10000
10000
15000
20000
25000
30000
35000
PSA 16 DAT (ET)

Fig S8.

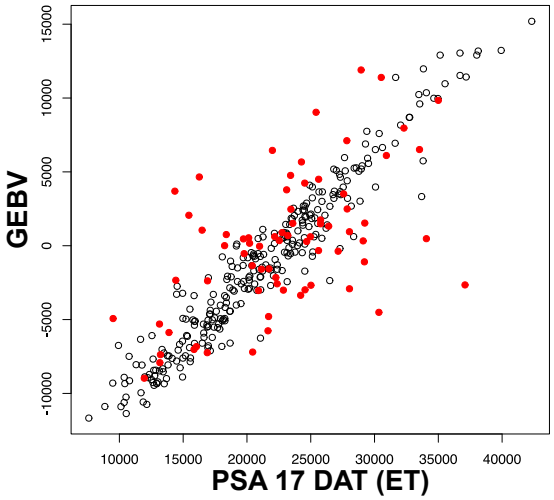

GEBV
15000
10000
5000
0
-5000
-10000
10000
15000
20000
25000
30000
35000
40000
PSA 17 DAT (ET)

Fig S9.

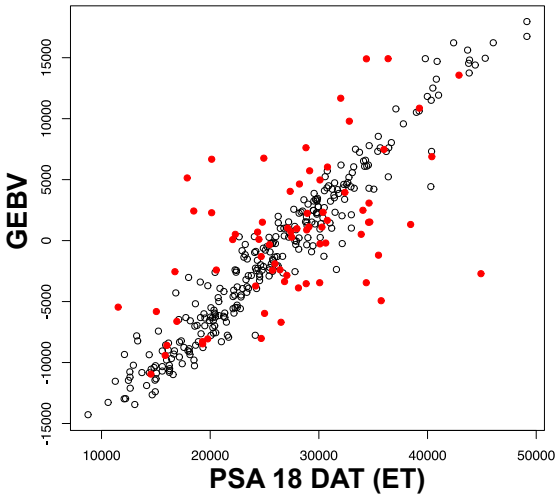


Fig S10.

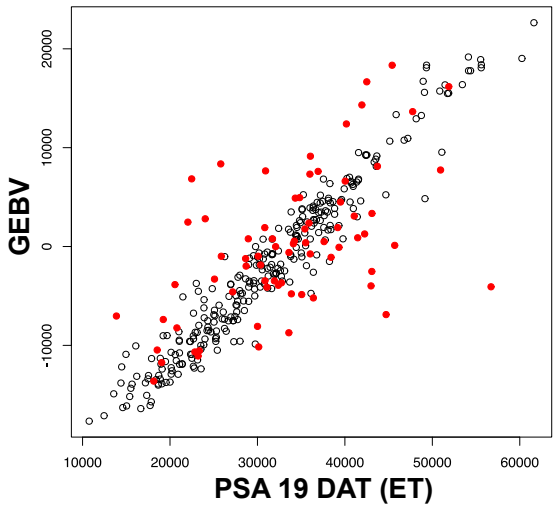


Fig S11.

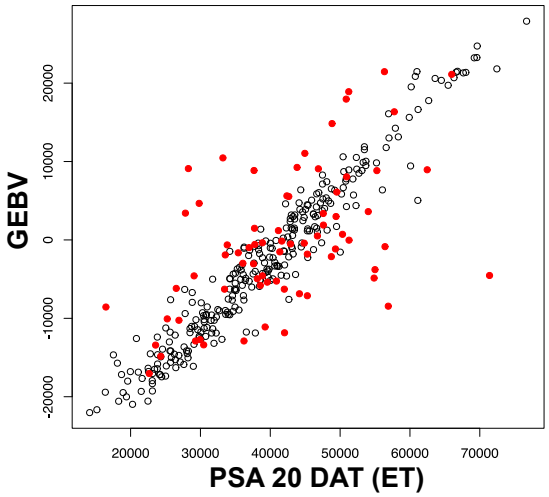


Fig S12.

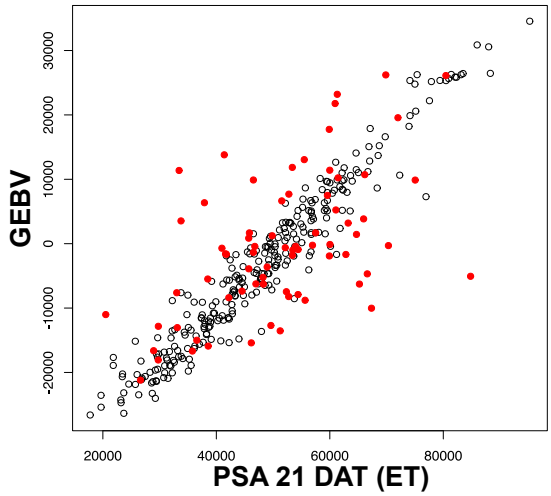


Fig S13.

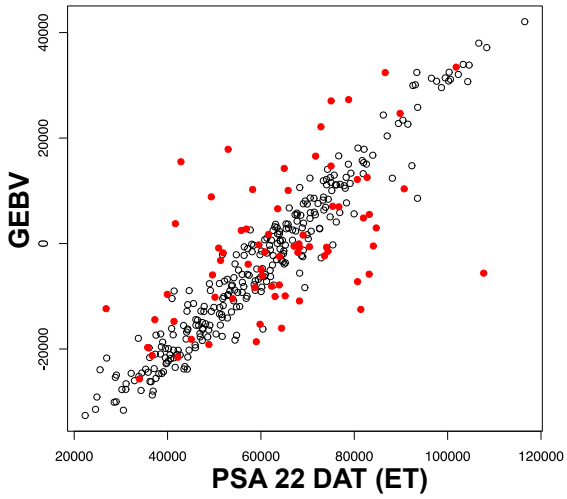


Fig S14.

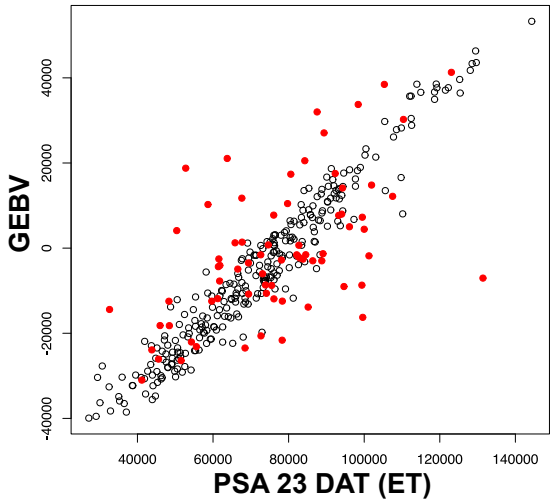


Fig S15.

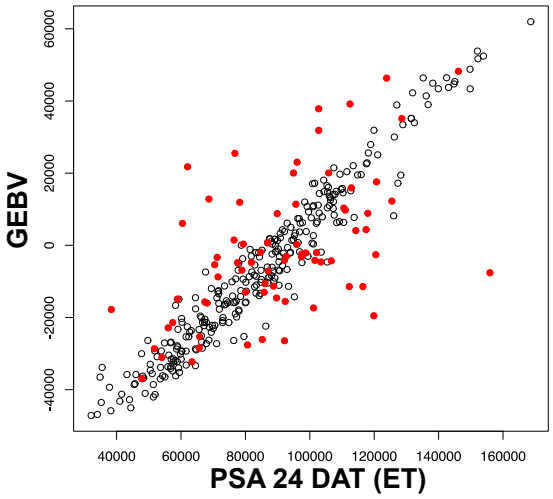


Fig S16.

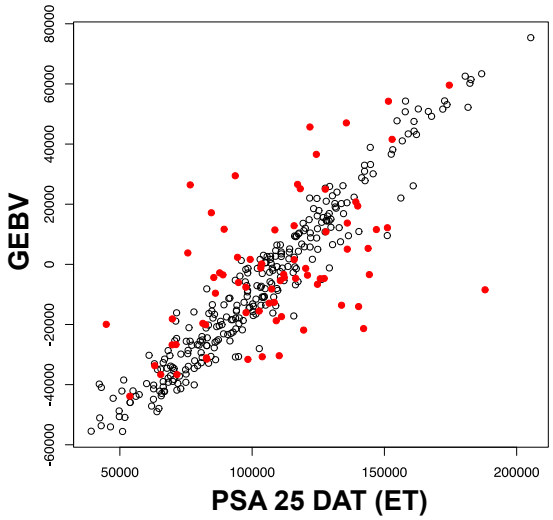


Fig S17.

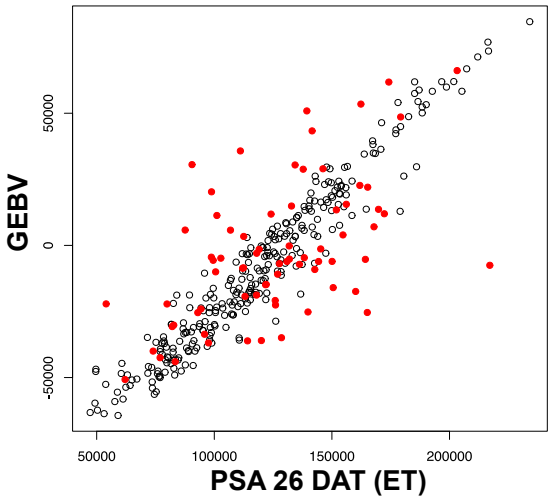


Fig S8.

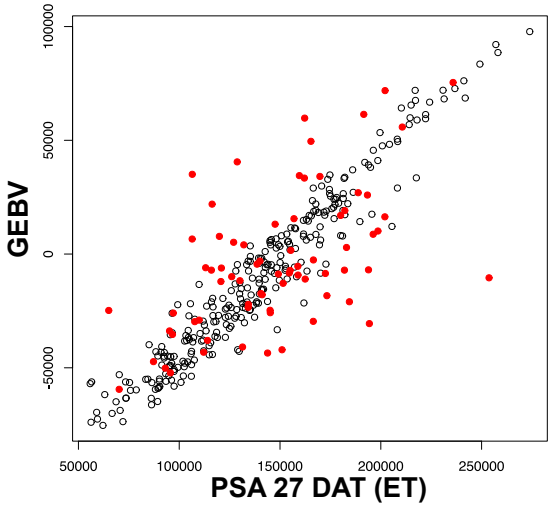


Fig S9.

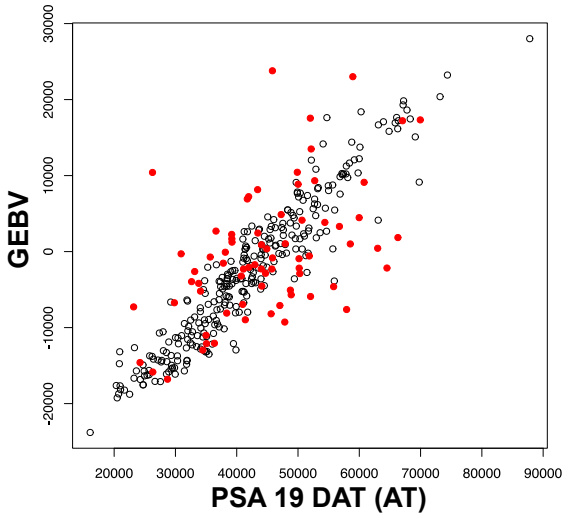


Fig S10.

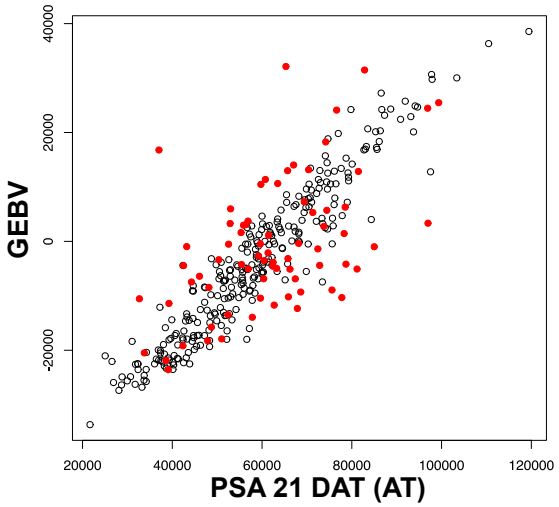


Fig S11.

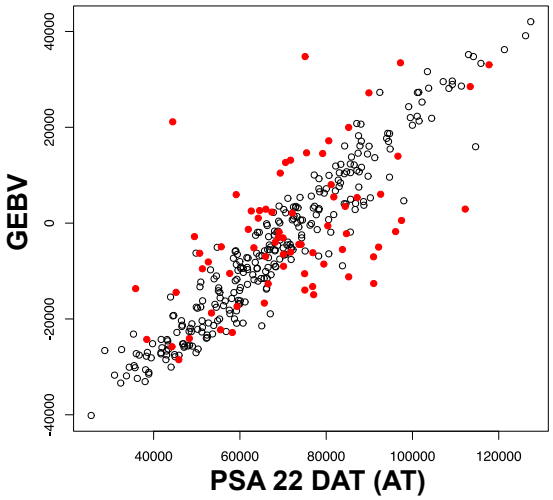


Fig S12.

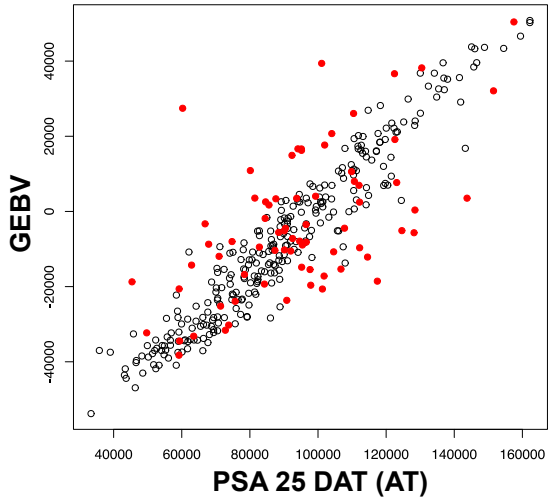


Fig S13.

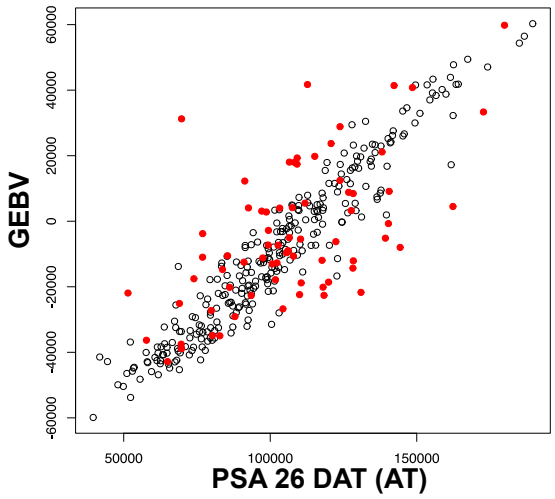


Fig S14.

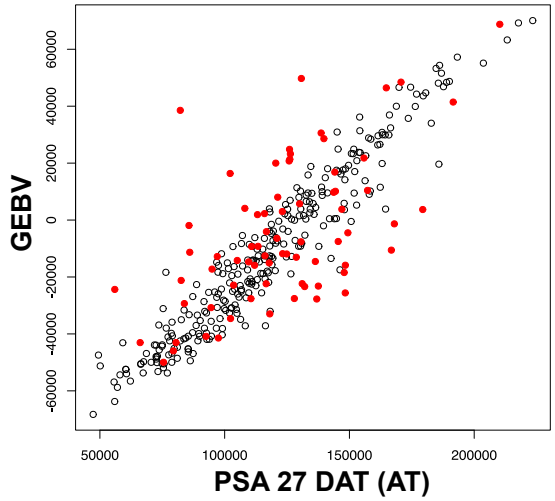


Fig S15.

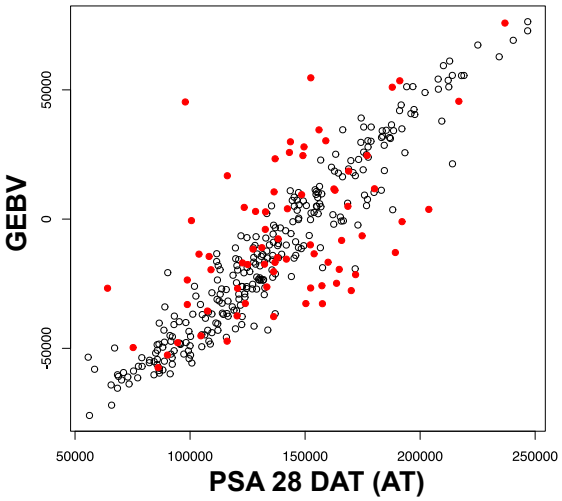


Fig S16.

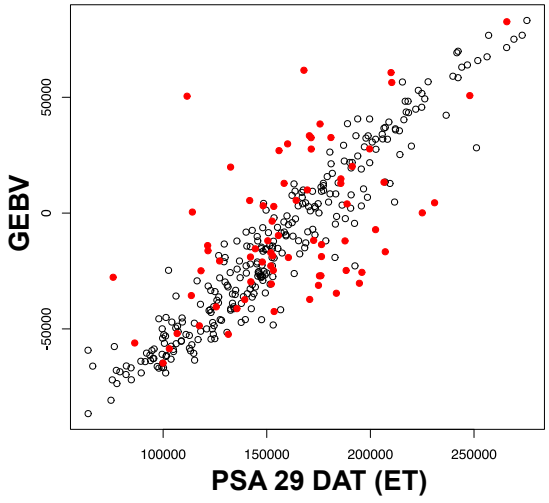


Fig S17.

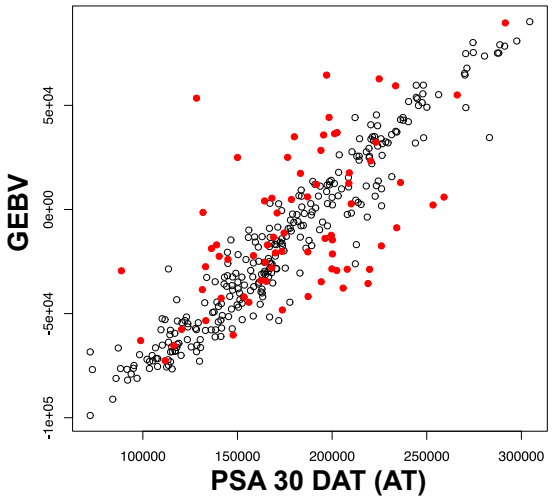


Fig S18.

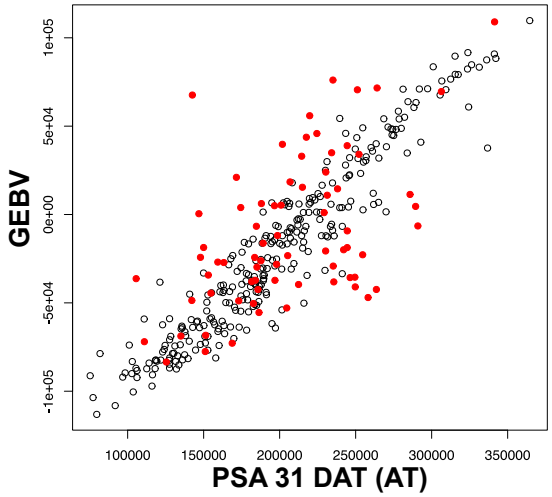


Fig S19.

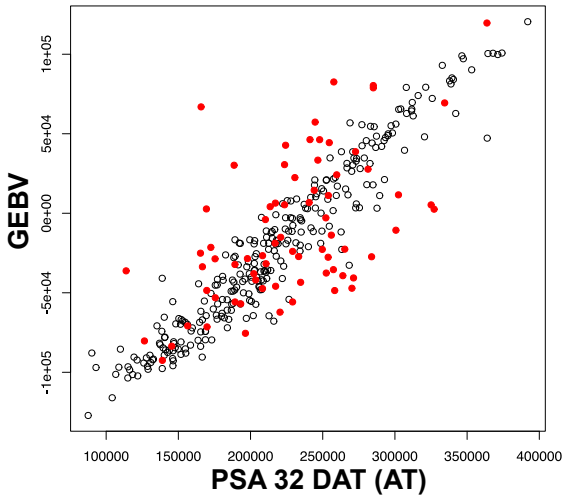


Fig S20.

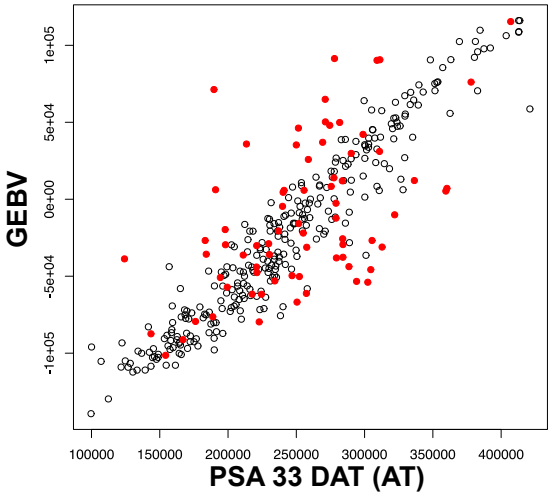


Fig S21.

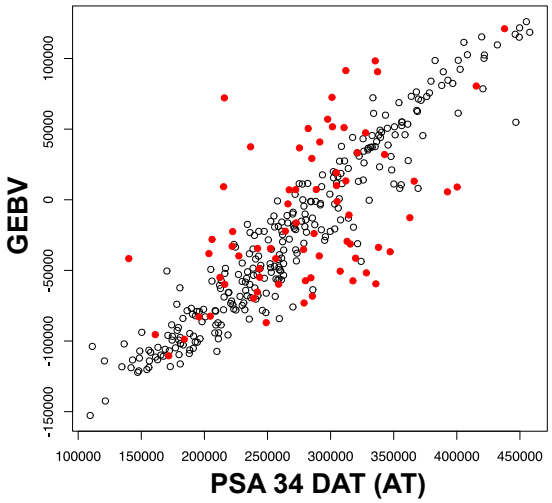


Fig S22.

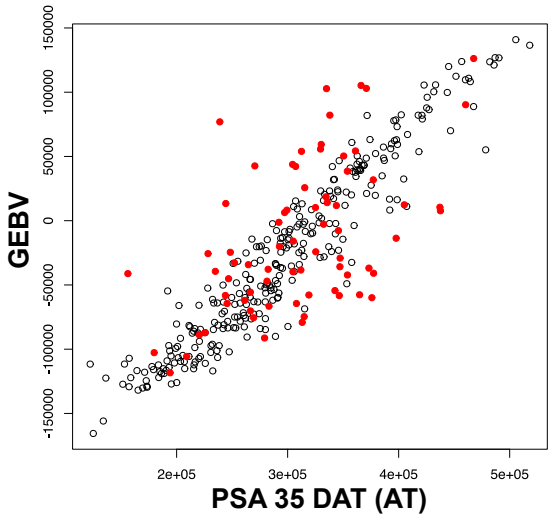


Fig S23.

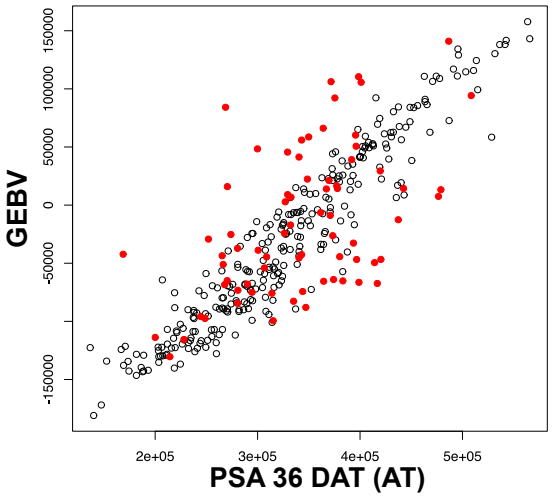


Fig S24.

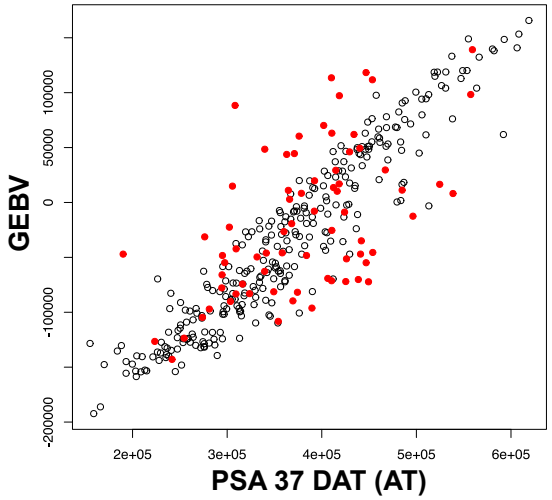


Fig S25.

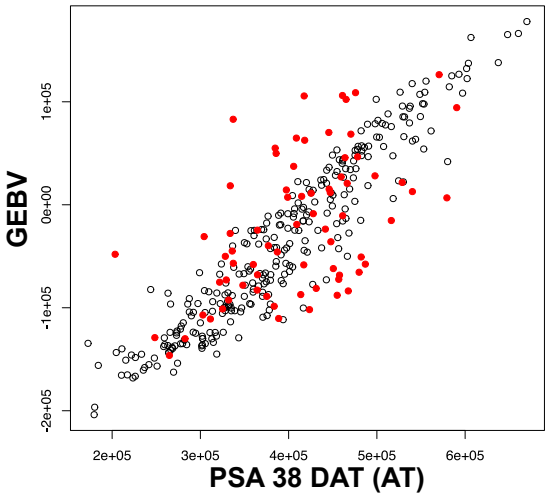


Fig S26.

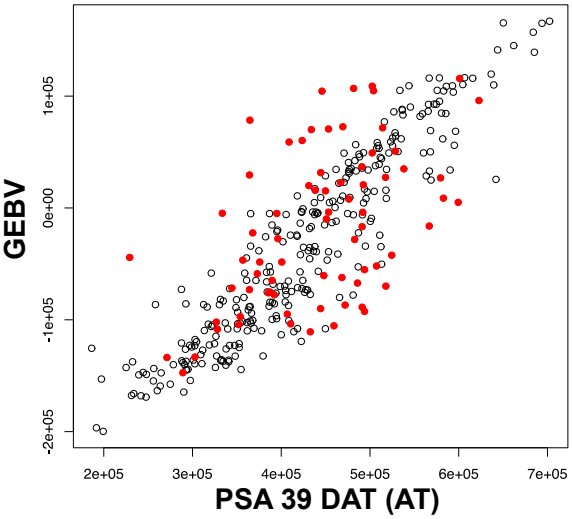


Fig S27.

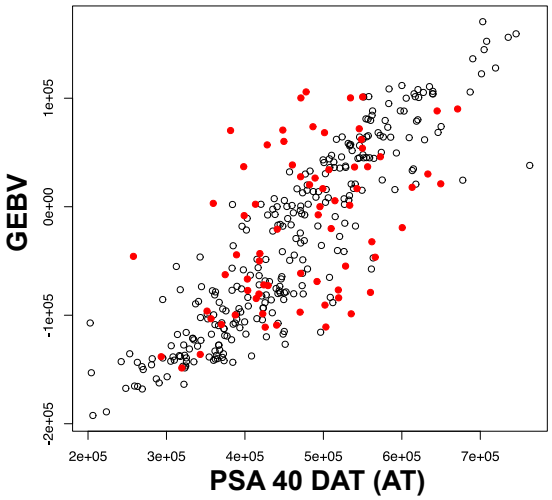
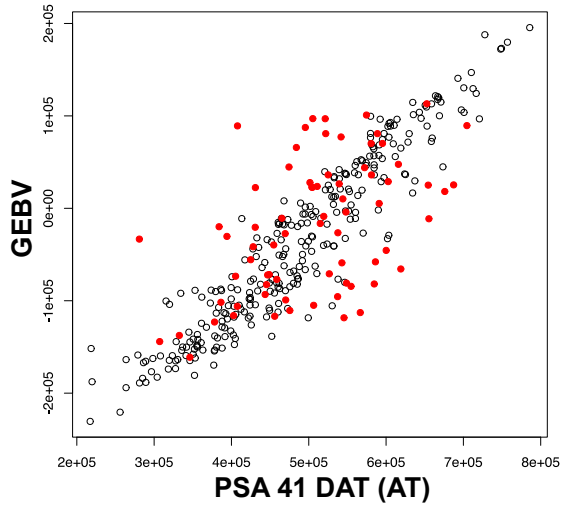


Fig S28.



Figures S1-S28. Comparison of genomic estimated breeding values (GEBV) and observed phenotypes for projected shoot area (PSA). GEBV were estimated using all 36,901 markers using a training population of 288 accessions. The black open points in each figure represent the accessions in the training population, while filled red points indicate accessions in the validation population (72 accessions). ET: early tillering; AT: active tillering

Table S1: Accessions used for RNA sequencing and/or GA quantification.

NSFTV ID	GSOR ID	IRGC ID	Accession Name	Country of origin	Sub-pop.	Experiment
NSFTV_109	301101	117817	MTU9	India	IND	RNA-seq
NSFTV_132	301123	117859	Rathuwee	Sri Lanka	IND	RNA-seq
NSFTV_157	301148	117908	Tainan Iku 487	Taiwan	TEJ	RNA-seq
NSFTV_16	301014	117658	Bico Branco	Brazil	AROMATIC	RNA-seq
NSFTV_162	301153	117915	TKM6	India	IND	RNA-seq
NSFTV_171	301162	117943	ZHE 733	China	IND	RNA-seq
NSFTV_187	301178	117606	C57-5043	United States	TRJ	RNA-seq
NSFTV_209	301200	117911	Tchibanga	Gabon	IND	RNA-seq
NSFTV_21	301019	117671	Byakkoku Y 5006 Seln	Australia	IND	RNA-seq
NSFTV_217	301208	117941	YRL-1	Australia	ADMIX	RNA-seq
NSFTV_25	301023	117676	Carolina Gold	United States	TRJ	RNA-seq
NSFTV_311	301301	117631	56-122-23	Thailand	TEJ	RNA-seq
NSFTV_33	301031	117688	Chuan 4	Taiwan	AUS	RNA-seq
NSFTV_347	301337	117695	Creole	Belize	TRJ	RNA-seq
NSFTV_36	301034	117609	CS-M3	United States	TEJ	RNA-seq; Horm. Quant.
NSFTV_363	301350	117730	Edomen Scented	Japan	TEJ	RNA-seq
NSFTV_37	301035	117698	Cuba 65	Cuba	TRJ	RNA-seq
NSFTV_39	301037	117700	NSF-TV 39	Bangladesh	ADMIX	RNA-seq
NSFTV_4	301004	117601	NSF-TV 4	India	AUS	RNA-seq
NSFTV_40	301038	117701	Dam	Thailand	ADMIX	RNA-seq
NSFTV_44	301041	117710	Dhala Shaitta	Bangladesh	AUS	RNA-seq
NSFTV_46	301043	117723	Dourado Agulha	Brazil	TRJ	RNA-seq
NSFTV_5	301005	117641	NSF-TV 5	India	AROMATIC	RNA-seq
NSFTV_51	301047	117727	Early Wataribune	Japan	TEJ	RNA-seq
NSFTV_55	301051	117611	Gerdeh	Iran	ADMIX	RNA-seq
NSFTV_57	301053	117739	NSF-TV 57	Iran	IND	RNA-seq
NSFTV_60	301056	117744	Gotak Gatik	Indonesia	ADMIX	RNA-seq
NSFTV_72	301065	117758	IR 8	Philippines	IND	RNA-seq
NSFTV_74	301067	117763	IRGA 409	Brazil	IND	RNA-seq
NSFTV_83	301075	117775	Kamenoo	Japan	TEJ	RNA-seq
NSFTV_9	301009	117647	Baber	India	TEJ	RNA-seq
NSFTV_380	301365	117909	Tainan-Iku No. 512	Taiwan	TEJ	Horm. Quant.
NSFTV_284	301274	117759	IR-44595	Nepal	IND	Horm. Quant.
NSFTV_389	301372	117689	CI 11011	United States	ADMIX	Horm. Quant.
NSFTV_214	301205	117626	WC 4419	Honduras	TRJ	Horm. Quant.
NSFTV_262	301252	117747	Halwa Gose Red	Iraq	AUS	Horm. Quant.
NSFTV_389	301372	117689	CI 11011	United States	ADMIX	Horm. Quant.
NSFTV_106	301098	117814	Ming Hui	China	IND	Horm. Quant.
NSFTV_222	301213	117842	Paraiba Chines Nova	Brazil	IND	Horm. Quant.
NSFTV_643	312009	117271	Minghui 63	China	IND	Horm. Quant.
NSFTV_11	301382	To be assigned	Baguamon 14	Bangladesh	IND	Horm. Quant.
NSFTV_86	301078	To be assigned	Kaw Luyoeng	Thailand	TEJ	Horm. Quant.
NSFTV_133	301124	117861	Rikuto Kemochi	Japan	TEJ	Horm. Quant.
NSFTV_307	301297	117925	Uzbekskii 2	Uzbekistan	TEJ	Horm. Quant.

Table S2: QTLs associated with PSA detected on overlapping days (19-27 days after transplant) in the early and active tillering experiments. DAT: days after transplant

Chr		QTL Start		QTL End		Early Tillering (DAT)		Active Tillering (DAT)
1		2036478		2236478				19, 21, 22, 25, 26, 27
2		34194749		34394749				26, 27
2		34516145		34716145		19, 21, 22, 25, 26, 27		25, 26, 27
4		5833594		6033594		19, 21, 22, 25, 26, 27		
4		19614956		19814956		26, 27		19, 21, 22, 25, 26, 27
5		4871496		5071496		21, 22, 25, 26, 27		19, 21
6		26135553		26335553		19, 22, 25, 26, 27		21, 22, 25, 26, 27
8		2901247		3101247				19, 21, 22, 25, 27
10		15297674		15497674		19, 21, 22, 25, 26, 27		19, 21, 22, 25, 26, 27
11		25751078		25951078				25, 26, 27

File S1: Phenotypic data for PSA and model parameters for all accessions of RDP1.

File S2: Pearson correlation coefficients for model parameters and PSA.

File S3: GWAS results for daily measurements of PSA. The values in each column represent the $-\log_{10}(p)$ value for each SNP.

File S4: Genes within 200kb of significant SNPs for GWAS of model parameters during the early and active tillering stages.

File S5: Read counts for the 2,417 genes found within 200 kb of a significant SNP from GWAS with model parameters.

File S6: Genes displaying significant differences in expression between allelic groups at id2010960 (FDR < 0.1).

File S7: Prediction accuracies for daily measurements of PSA during the early and active tillering stages.

CHAPTER 5

INTEGRATING IMAGE-BASED PHENOMICS AND ASSOCIATION ANALYSIS TO DISSECT THE GENETIC ARCHITECTURE OF TEMPORAL SALINITY RESPONSES IN RICE

Malachy T. Campbell, Avi C. Knecht, Bettina Berger, Chris J. Brien, Dong Wang, Harkamal Walia

Published: June 25 2015; doi: http://dx.doi.org/10.1104/pp.15.00450

Abstract

Salinity affects a significant portion of arable land and is particularly detrimental for irrigated agriculture, which provides a third of the global food supply. Rice, the most important food crop is salt-sensitive. The genetic resources for salt tolerance in rice germplasm exist but are under-utilized due to the difficulty in capturing the dynamic nature of physiological responses to salt stress. The genetic basis of these physiological responses is predicted to be polygenic. In an effort to address this challenge, we generated temporal imaging data from 378 diverse rice genotypes across 14 days of 90 mM NaCl stress and developed a statistical model to assess the genetic architecture of dynamic salinity-induced growth responses in rice germplasm. A genomic region on Chromosome 3 strongly associated with the early growth response and was captured using visible range imaging. Fluorescence imaging identified four genomic regions linked to salinity-induced fluorescence responses. A region on chromosome 1 regulates both the fluorescence shift indicative of the longer-term ionic stress and the early growth rate decline during salinity stress. We present a new approach to capture the dynamic plant responses to its

environment and elucidate the genetic basis of these responses using a longitudinal genome-wide association model.

5.1 Introduction

Nearly one-third of the 54 million ha of the highly saline soils in the world are located in South and South-east Asia. Rice (*Oryza sativa*), which is the primary source of calories and protein for these two regions, is very sensitive to salinity stress with even moderate salinity levels known to decrease yields by 50% (Zeng et al., 2002). Projected sea level rise due to climate change is expected to increase salt-water ingress in coastal rice-growing regions of South and Southeast Asia. Therefore, development of salt-tolerant rice cultivars is essential to maintain rice productivity in the salinity-affected regions globally.

Salt tolerance, defined as the ability to maintain growth and productivity in saline conditions, is a complex polygenic trait that may be influenced by distinct physiological mechanisms (for a comprehensive review of genes involved with salinity tolerance in rice see Negrao et al 2011 (Munns et al., 1982; Munns and Termaat, 1986; Cheeseman, 1988; Munns and Tester, 2008; Negrão et al., 2011; Horie et al., 2012) At the cellular level, plants respond to saline conditions in two phases, namely an osmotic (shoot ion independent) and an ionic stress phase, which can occur in an overlapping manner with varying intensity during the course of salinity stress (Munns and Termaat, 1986; Munns, 2002; Munns and James, 2003; Munns and Tester, 2008; Horie et al., 2012). During the osmotic stress phase, which occurs soon after the onset of salinity, the reduction in external osmotic potential disrupts water uptake and impedes cell expansion, which, at the whole plant level, leads to reduced growth rate (Matsuda and Riazi, 1981; Munns and Passioura, 1984; Rawson and Munns, 1984; Azaizeh and Steudle, 1991; Fricke and Peters, 2002; Fricke, 2004; Boursiac et al., 2005). As salinity stress persists over several days and

weeks, sodium ions (Na^+) accumulate to toxic levels, resulting in cell death and precocious leaf senescence (Lutts et al., 1996; Munns, 2002; Munns and James, 2003; Ghanem et al., 2008). This is typically observed during the ionic phase of the salinity response (Munns, 2002; Munns and James, 2003; Munns and Tester, 2008). Plants possess distinct mechanisms to adapt to these osmotic and ionic stresses that are controlled by a suite of genes (Apse et al., 1999; Carvajal et al., 1999; Ishitani et al., 2000; Shi et al., 2000; Zeng and Shannon, 2000; Halfter et al., 2000; Rus et al., 2001; Martínez-Ballesta et al., 2003; Berthomieu et al., 2003; Boursiac et al., 2005, 2008; Ren et al., 2005; Huang and Spielmeyer, 2006; Obata et al., 2007; Davenport et al., 2007; Székely et al., 2008; Horie et al., 2011; Rivandi et al., 2011; Asano et al., 2012; Munns et al., 2012; Schmidt et al., 2013; Latz et al., 2013; Liu et al., 2014; Campo et al., 2014; Choi et al., 2014). The genetic basis of temporal adaptive responses to salinity stress remains to be explored in rice and other crops. This is primarily due to challenges in capturing the dynamic physiological responses to salinity for a large number of genotypes in a non-destructive manner. Manual phenotyping to detect incremental changes in growth rate during the osmotic stress and slight shifts in leaf color due to ionic stress are difficult to quantify for a large number of genotypes.

In rice, at least one major QTL (*saltol*) for salinity tolerance has been characterized based on end point measurements of biomass, senescence/injury and Na^+ and K^+ concentrations (Bonilla et al., 2002; Lin et al., 2004; Thomson et al., 2010). *SKC1* is the causative gene underlying the *saltol* region. *SKC1* encodes a Na^+-selective high affinity potassium transporter (*HKT*) that regulates Na^+/K^+ homeostasis during salinity stress (Ren et al., 2005). High levels of Na^+ displace cellular K^+, an essential element for several enzymatic reactions and physiological processes (Gierth and Mäser, 2007). The ability to maintain cellular K^+ levels during salinity through the action of Na^+-

selective potassium transporters or Na^+/H^+ antiporters is a well-characterized tolerance mechanism in cereals including rice (Ren et al., 2005; Sunarpi et al., 2005; Huang and Spielmeyer, 2006; Møller et al., 2009; Mian et al., 2011; Munns et al., 2012).

Numerous studies have utilized conventional linkage mapping to identify QTLs for morphological and physiological responses to salinity in rice using discrete endpoint measurements (Bonilla et al., 2002; Lin et al., 2004; Ren et al., 2005; Negrão et al., 2011; Wang et al., 2012). However, the physiological adaptation to saline conditions is a complex continuous process that develops over time. While some accessions will exhibit similar end-point phenotypic values, the genetic and physiological mechanisms giving rise to the similar phenotypes may be very different and the growth trajectories throughout the experiment may be distinct. Although single time point studies have yielded important information regarding the genetic basis of salinity tolerance, such approaches are too simple to reveal the genetic architecture of stress adaptation. With the advent of high-throughput image-based phenotyping platforms, it is now feasible to quantify dynamic responses during the stress treatment for a large number of genotypes (Berger et al., 2010; Golzarian et al., 2011; Honsdorf et al., 2014; Chen et al., 2014).

Image-based phenotyping has been combined with GWAS and linkage mapping to examine the genetic basis of complex developmental processes (Busemeyer et al., 2013; Topp et al., 2013; Moore et al., 2013; Slovak et al., 2014; Würschum et al., 2014; Yang et al., 2014; Bac-Molenaar et al., 2015). Moreover, the introduction of the time axis provides a better understanding of the physiological processes underlying complex stress and developmental responses compared to single end-point measurements (Zhang et al., 2012; Moore et al., 2013; Brown et al., 2014; Slovak et al., 2014; Chen et al., 2014; Bac-Molenaar et al., 2015). However, to date no studies

have implemented an association mapping approach using image-derived phenotypes to address the genetic basis of dynamic stress responses in plants. Image-based phenotyping offers several advantage over conventional phenotyping: (1) quantitative measurements can be recorded over discrete time points to capture morphological and physiological responses in a non-destructive manner, (2) the use of various types of spectral imaging to address phenotypes that are not detectable to the human eye such as chlorophyll fluorescence and leaf water content. Integrating dynamic phenotypic data and association mapping has the potential to query genetic diversity across hundreds of accessions for complex traits, and provides much higher resolution compared to conventional linkage mapping. Here, we explored the dynamic growth and chlorophyll responses to salinity of a diverse set of rice accessions using high-throughput visible and fluorescence imaging. To assess the genetic basis of plant growth in saline conditions a logistic model was used to describe the temporal growth responses and was incorporated into the statistical framework necessary for association mapping. Coupled with temporal fluorescence imaging, we present new insights into the genetic architecture of osmotic and ionic responses during salinity stress in rice.

5.2 Results

5.2.1 Capturing osmotic and ionic components of salinity stress using high-throughput imaging

To assess the ionic and osmotic components of salinity response in rice, a diversity panel consisting of 373 rice lines was exposed to 90 mM NaCl during the early tillering stage (Supplemental Fig. S1A, Fig. S2). Morphological and physiological responses were monitored over a period of 14 days after 90 mM NaCl treatment using non-destructive imaging (Supplemental Fig. S1B). Two types of imaging systems, RGB and FLUO, were utilized to address osmotic and ionic components of salinity stress respectively. RGB/Visible imaging

allows growth and other morphological parameters to be quantified in a non-destructive manner, and has been used to monitor growth shortly (1-10 days) after the onset of salinity to address primarily the shoot ion-independent phase (osmotic phase) of salinity stress (Rajendran et al., 2009). The effects of salinity on chlorophyll and other fluorophores can be monitored by detecting changes in the color or intensity of pixels, thus providing important information regarding the ionic component of salinity stress (Berger et al., 2010).

We developed an open-source image processing software named "Image Harvest" to extract plant pixels from images and quantify fluorescence color ranges (Materials and Methods; Supplemental Fig. S1C). Image Harvest is publically available for download and is optimized for processing high volume plant image data using parallel computing. The dataset generated for the current study consisted of 142,671 RGB images and 95,118 fluorescence images, which were processed using LemnaGrid and ImageHarvest, respectively. The entire dataset is available on iPlant (http://mirrors.iplantcollaborative.org/browse/iplant/home/shared/walia_rice_salt). The resulting output consists of temporal data for 97 digital (pixel-based) traits, seven of which are used to describe plant morphology and growth traits, while the remaining 90 traits are used to describe fluorescence responses (Supplemental Table S1).

5.2.2 Assessing salinity-induced growth responses in the rice diversity panel

To determine how accurately the biomass-related, image-derived growth metrics represented actual biomass and total shoot area, plants from 72 accessions were harvested on the last day of the imaging experiment and fresh and dry mass was recorded and total plant area was directly measured using a leaf area meter (LICOR, LI-3100C). Correlation analysis was done using seven biomass-related metrics and three manual measurements. Of the seven biomass-related metrics derived from color images, Projected Shoot Area (PSA), which is defined at the sum of all pixels

from all three RGB images, showed the highest correlation with all manual biomass-related measurements (Supplemental Table S2). As expected, PSA showed the strongest positive correlation with total plant area also (r^2=0.96, p<0.001, n=72; Fig. 1A). Shoot fresh and dry weight showed a strong positive correlation with PSA, although at a slightly lower correlation compared to total plant area (r^2=0.95, p<0.001 for both shoot fresh and dry weight; Figs. 1B and 1C, respectively). A significant difference in PSA was detected between treatments using a one-way blocked ANOVA (where accession is considered as a block) beginning at day 4 after 90 mM NaCl (p < 0.0028; Fig. 1D). These results indicate that PSA is an accurate and sensitive metric for assessing plant biomass accumulation in response to salinity.

To determine whether there were any differences in the salinity response among the five subpopulations as classified by Zhao et al. (Zhao et al., 2011), we calculated the salinity-induced growth response as the ratio of PSA in salt treated plants over control plants. *Aromatic* lines were excluded due to small number of accessions. For each subpopulation, the salinity-induced growth response was modeled across all time points with a decreasing logistic curve. Therefore, on day 1 of salt treatment the growth-response is 1, it begins to decrease after the onset of salinity stress and eventually flattens out as vegetative growth declines and plants transition to reproductive phase. Pairwise comparisons of growth-response models revealed significant differences between several subpopulations (Table I). Notably, the *tropical japonica* subpopulation showed a significantly lower growth reduction in response to salinity when compared to other subpopulations, suggesting that this varietal group may be an important source for osmotic stress tolerance response during early stages of salinity stress (Fig. 1E). The *admix* and *aus* subpopulations showed the most severe reduction in growth. *Aus* accessions displayed the earliest reduction in PSA, with a significant difference observed four days after 90mM salt

application ($p < 6.18 \times 10^{-5}$). While for the *Admix* subpopulation, significant differences between treatments were observed beginning seven days after the onset of 90mM salinity ($p < 1.31 \times 10^{-6}$).

5.2.3 Assessing salinity-induced chlorophyll responses

To assess the effects of salinity stress on leaf senescence, plants were imaged in a separate fluorescence-imaging chamber. Since the available functions are limited in LemnaGrid software, we developed an open source processing software called ImageHarvest to extract several spectral metrics from the 95,118 fluorescence images. Color ranges that may be indicative of salinity-induced chlorophyll responses were identified by utilizing an ad hoc image segmentation strategy that classified the range of colors present in all fluorescence images into 90 classes of color ranges. Based on our pixel classification strategy, we identified thirty-two color classes that showed significant differences between treatments after three or more days of salt stress across all 373 accessions ($p < 0.00056$; Fig. 2A). No difference between treatments was observed for any of the 32 color classes before the application of NaCl. Canonical correlation analysis between pixel classes within each time-point showed strong correlations between color classes (Supplemental File S1; Supplemental File S2). As senescence progresses the color properties of the fluorescence signal emitted from stressed tissue will change over time. Since our ad hoc segmentation approach classifies pixels into discrete color classes, it is likely that over time pixels that represent stressed tissue will change membership between classes with similar color properties. Therefore to examine the relationship between color classes across time points we performed hierarchal clustering analysis (HCA) using the mean temporal trend for each color class across all 373 accessions for each treatment. No clear distinction between treatments could be observed from hierarchal clustering, rather clustering seemed to be driven largely by the

behavior of color classes over time (Fig. 2B). If fluorescence signals were influenced only by salinity stress, one would expect clusters to be formed based on treatment. These results suggest that developmental processes likely influence the fluorescence signals. However the significant differences observed between treatments suggest that the onset of salinity may affect the timing or magnitude of these signals.

Significant differences between treatments were generally observed during the later time points, with the majority of color classes displaying significant differences after day 11 of 90 mM NaCl (Fig. 2A). However, color classes that displayed significant differences during the early stages of salinity stress tended to populate cluster 1. For instance at seven days after 90 mM NaCl five of the seven color classes that displayed significant differences between treatments were members of cluster 1. This suggests that fluorescence responses represented by cluster 1 may be important digital markers for monitoring the early effects of salinity stress on chlorophyll responses in rice.

5.2.4 Development of model for association analysis of longitudinal salinity-induced growth responses

Since PSA was most strongly correlated with conventional biomass related measurements and displayed significant differences between treatments during the early phase of salt treatment, we sought to identify genetic loci associated with the salinity-induced growth response (defined here as the square root transformed ratio of the projected shoot area (PSA) of plants under salt stress over that of plants in control conditions) of 360 rice accessions for which phenotypic and genetic information was available (Zhao et al., 2011; Famoso et al., 2011). Briefly, for each accession and SNP, the growth response over 18 days is fitted with a decreasing logistic curve for each genotype. To perform statistical inference, the kinship matrix between lines is used to account for the population structure of the diversity panel and a first order autoregressive covariance

structure is used to account for the dependency between time points. For each SNP, a likelihood ratio test is carried out on whether the growth response curves are the same for the two genotypes. The model accounts for both genetic relationships between accessions due to rice subpopulation structure and the inherent non-independent nature of daily observations.

The longitudinal growth response model was compared to a conventional mixed model in which the response ratio at each individual time-point was used as the phenotypic variable. While several peaks were identified with both methods we observed a considerably higher number of significant QTL and considerable lower *p*-values with the longitudinal model. For instance, a total of 115 highly significant SNPs, which corresponds to 55 QTL, were identified when the longitudinal model was used ($p < 10^{-8}$; Fig. 3A; Supplemental File S3). In contrast, the maximum number of significant QTL identified at only one time-point was 26 when the segmented model was used ($p < 10^{-4}$; Fig. 3A; Supplemental File S4). Though the number of significant SNPs and the number of loci are not proportional, these results indicate that the collective analysis of multiple time points significantly improved the ability to identify loci associated with complex polygenic traits such as salinity-induced growth responses.

5.2.5 Genetic architecture of salinity-induced temporal growth response

Our genetic association analysis using the longitudinal model identified several highly significant peaks in close proximity to genes with putative roles in salt tolerance and or ion homeostasis in rice and other species based on published literature. A highly significant cluster of SNPs was located at approximately 16.3 Mb on chromosome 3 (id3008139; $p < 1.11 \times 10^{-16}$; Fig. 3B). This region was identified with both the longitudinal growth response and mixed model GWA methods. The large peak on chromosome 3 is populated by ten highly significant SNPs converging on a ~240 Kb region, and includes 89 gene models based on the MSU Rice

Annotation Project Release 7 (Supplemental File S3). We found a gene within a 200 kB window surrounding the most significant SNP (id3008139) on chromosome 3 that encodes a putative Shaker family inward rectifying potassium channel (*POTASSIUM TRANSPORTER1*; LOC_Os03g28120) gene (Table S3). These K^+ channels mediate cellular potassium uptake and have been shown to increase salinity tolerance via lowered Na^+:K^+ ratios in yeast (*Saccharomyces cerevisiae*) and rice (Obata et al., 2007). The accessions with the minor allele at this locus displayed a greater reduction in plant growth in response to saline conditions when compared to the major allele genotypes (Fig. 3C). The minor allele was represented at a greater proportion in the *temperate japonica*, *aus* and *admix* subpopulations with approximately 77, 36 and 46% of accessions retaining the allele respectively, suggesting that salt sensitivity associated with this locus is present in multiple subpopulations.

On chromosome 1, a cluster of highly significant SNPs was detected around approximately 25 Mb (id1014913; $p < 3.18 \times 10^{-11}$). The minor allele at SNP id1014913 was significantly under-represented in the *japonica* varietal groups, with no *tropical japonica* accessions and only a single *temperate japonica* variety possessing the minor allele, suggesting that this locus may be fixed between *indica* and *japonica* subspecies. In general, the accessions with the minor allele at id1014913 displayed a greater reduction in growth in response to salinity when compared to the major allele group, with a slight difference observed as early as 4 days after reaching the 90 mM NaCl and progressively greater difference during later time points (Fig. 3D). GWAS conducted with leaf Na^+ content, showed a minor peak at approximately 30 Mb, suggesting that this 5 Mb region on chromosome 1 may influence both growth as well as Na^+ ion homeostasis (id1018154, 30.2 Mb, $p < 8.44 \times 10^{-5}$; Supplemental Fig. S3A). Varieties with the major allele at id1014913 had higher Na^+ content and Na^+:K^+ ratio, as well as lower K^+ content ($p < 0.001$; Supplemental

Fig. S4A and S4B). These results suggest that the mechanism underlying the more tolerant response observed in the major allelic group might be independent of ion homeostasis, or accession from this allelic group may be able to tolerate higher cellular Na^+ content without detrimental effects on growth.

5.2.6 Genetic architecture of salinity-induced chlorophyll responses

To identify loci associated with salinity-induced chlorophyll responses, association mapping was performed at each time point individually using a conventional mixed model for each of the 32 salinity-responsive color classes (those with $p < 0.00056$). For each color class and accession we calculated the salinity response, which is defined as the percent of total pixels from two side-view images accounted for by the color class in saline conditions minus the percent of total pixels accounted by the class in control conditions. In total, four classes showed significant associations at one or more time points throughout the experiment ($p < 1.0 \times 10^{-7}$; Fig. 4A; Supplemental Fig. S5-18; Supplemental File S5). Highly significant signals were detected on chromosomes 1, 3, 6 and 10 for color classes 21, 31, 41 and 52 at multiple time points. HCA of color classes in control and saline conditions revealed strong similarities in the temporal behavior of color classes 31, 41 and 52 (Supplemental Fig. S19 and S20). Interestingly, HCA showed a clear separation between treatments for these color classes (Fig. 2B). The salinity response of classes 31, 41 and 52 were tightly grouped in cluster 5, while classes 41 and 52 displayed similar behavior in control conditions. These results indicate that these fluorescence-based metrics are measuring similar phenotypic responses, which may explain the overlapping genomic regions associated with these metrics.

The most significant association was observed on day 14 after 90 mM NaCl for class 41 ($p < 1.14 \times 10^{-14}$; id1015984). The corresponding SNP was located at approximately 27.6 Mb on

chromosome 1. Notably, this region was also associated with several other metrics at later time points during the experiment. For instance, the earliest significant signal for class 21 localized to this same region on chromosome 1 and was detected only on day 2 after 90 mM NaCl. However, significant associations were not observed again in this region until day 5 after 90 Mm NaCl (class 52; $p < 9.91 \times 10^{-9}$; id1015991).

5.3 Discussion

Excess soil sodium reduces plant growth through osmotic and ionic effects. Shortly after exposure to excess Na^+, (before Na^+ accumulates in the cell), the osmotic potential outside the cell is reduced which impairs cell expansion at the cellular level and disrupts plant-water relations, carbon assimilation and transpiration at the plant level. In time, Na^+ accumulates in the cytoplasm to toxic levels, resulting in cell death and senescence. The temporal nature of excess Na^+ on plant growth adds an additional level of complexity and complicates the study of physiological mechanisms conferring tolerance and the underlying genetic basis of these polygenic traits.

When quantifying these dynamic processes, the choice of a sampling time point for end-point measurements is critical for detecting variance for the trait of interest. When evaluating large diversity panels for a given trait, determining the appropriate time point to sample is very difficult, since such germplasm collections are designed to capture a large portion of the genetic diversity in a species and should exhibit a wide range of phenotypic values. Therefore, sampling at only a single time point for the trait of interest may yield a significant underrepresentation of the variance for the trait in the population. The quantification of the trait across many time points, which in some instances may be laborious and unfeasible, eliminates many of the pitfalls associated with single end point measurements and adds another dimension that may help dissect

and further elucidate the complex biological processes underlying the phenotype. However, with the advent of high-throughput image-based phenomics platforms the quantification of morphological and physiological traits across multiple time points is expected to become routine for many laboratories.

Recent studies by Moore et al. (2013), Yang et al. (2014), and Wurschum et al. (2014), utilized image-based phenomics to identify QTLs involved with temporal developmental dynamics in Arabidopsis, rice and wheat, respectively (Moore et al., 2013; Würschum et al., 2014; Yang et al., 2014). Genetic analysis conducted at discrete time points identified several time-specific QTLs that partially influence the final phenotype, but would only be detected at a specific time point. In the current study, a similar univariate association analysis was performed at each time point for the 32 fluorescence color classes. The significant signals identified from fluorescence imaging displayed a high degree of temporal dynamics. For instance, a region localized to ~27.6 Mb on chromosome 1 displayed significant associations in both the early and later time points of the experiment, with the earliest significant signal in this region being detected on day 2 after 90 mM NaCl for class 21, but were not observed again in this region until day 5 (class 52; 9.91×10^{-9}; id1015991). While, on the final day of the experiment significant associations were detected only for classes 41 and 31. The presence of these signals at discontinuous intervals across time points suggest that these loci may influence the processes that give rise to the fluorescence phenotype, and would have been missed if fluorescence responses were measured at a single time point. The inclusion of multiple time points provides additional insight into the genetic architecture underlying dynamic fluorescence responses.

The presence of favorable alleles in close proximity provides an opportunity to utilize this region for breeding programs. Several significant associations were identified for growth response,

chlorophyll health and leaf Na^+ content in a region spanning ~5 Mb on chromosome 1 (~25-30 Mb). This overlap among multiple traits is supported by canonical correlation analysis of multiple traits derived from image-based and conventional phenotyping (Supplemental File S2). Class 41 showed a strong negative correlation (r^2=-0.50 at day 12 after 90 mM salt application) with growth (PSA) reduction during salinity, indicating that lines exhibiting a more tolerant growth response also exhibited reduced senescence (Supplemental File S2). No significant/strong correlations were observed for image-based traits and leaf ion content. Our results suggest that this 5 Mb region on chromosome 1 may influence both growth and senescence during salinity, and may serve as an important source of genetic diversity for development of salt-tolerant rice varieties that are able to maintain growth and minimize senescence in saline conditions. Interestingly, several other studies have identified QTL associated with seedling survival, Na^+ uptake and Na^+:K^+ ratio, chlorophyll content and ion homeostasis spanning this region on chromosome 1 (Koyama et al., 2001; Lin et al., 2004; Thomson et al., 2010). These studies were conducted during the early vegetative growth stage of rice using a similar NaCl concentration as the present study, and support the conclusion that these QTL may regulate aspects of salinity tolerance during the early tillering stage in rice. While plants have distinct mechanisms to respond at the cellular level to the osmotic and ionic effects of salinity, salt tolerance, defined as the ability to maintain growth in saline conditions, may be the result of tolerant ionic or osmotic responses, or a combination of both. However, dissecting the precise physiological response mechanisms (osmotic or ionic) underlying these loci requires further experimentation.

Functional data analysis allows complex longitudinal phenotypic data to be reduced to a single mathematic function with a limited set of parameters that fully capture and describe the

dynamics of biological processes (Paine et al., 2012). This approach simplifies phenotypic comparisons between individuals across multiple time points and can be used to accurately predict future outcomes. Time-series datasets derived from image-based phenotyping has been combined with functional data analysis to examine shoot and root growth in response to various environmental conditions (Walter et al., 2002; van der Weele et al., 2003; Poiré et al., 2014; Chen et al., 2014; Bac-Molenaar et al., 2015). Functional data analysis can be combined with conventional genetic analysis such as linkage mapping and GWAS to identify loci that may be regulating dynamic processes (Cui et al., 2006; Wu and Lin, 2006; He et al., 2010; Das et al., 2011; Bac-Molenaar et al., 2015).

In our experiments, the temporal and polygenic nature of plant salinity responses required the development of novel statistical approaches that combine information across time points to identify loci with minor effects that regulate the adaptation of growth to saline conditions. In this study we have leveraged novel image processing software and longitudinal GWA methods to examine the genetic architecture of salinity-induced growth responses in rice. The salinity-induced growth response GWA model greatly increased the ability to detect SNPs with minor effects compared to the conventional mixed model approach. A highly significant cluster of SNPs was located at approximately 16.3 Mb on chromosome 3 (id3008139; $p = 1.11 \times 10^{-16}$), and was identified with both the longitudinal growth response and mixed model GWA methods. However, additional highly significant signals that impacted growth, senescence and Na^+ content were detected only with the longitudinal model. Similar power gains for detecting genetic associations from longitudinal phenotypes have been reported by Xu et al. (2014) and Wu and Lin (2006) (Wu and Lin, 2006; Xu et al., 2014). However, to date, no studies have leveraged longitudinal genetic algorithms for large-scale association studies of image-based data from

plants. Our results suggest that longitudinal phenotypes can provide an advantage over typical cross-sectional (i.e. endpoint) phenotypic data for studying the genetic architecture of complex abiotic stress responses in plants. This study highlights the potential of these new technologies and provides a framework for future studies to integrate next-generation phenotyping technology with association mapping to understand the genetic architecture of complex polygenic traits such as salinity tolerance.

5.4 Materials and Methods

5.4.1 Greenhouse Conditions and salt treatment

Seeds from 373 genotypes from the rice diversity panel were surface sterilized with fungicide, Thiram® and germinated on moist paper towels in plastic boxes for three days (Zhao et al., 2011; Famoso et al., 2011). Three uniformly germinated seeds of each genotype were transplanted to pots (150mm diameter x 200 mm height) filled with 2.6 kg of UC Mix and placed into square containers to allow for water to collect. Plants were thinned to one seedling per pot six days after transplanting (DAT). For the first 7 DAT each pot was watered daily with ~150 mL from the top of the pot. Over the course of the three experiments, the greenhouse temperatures during the day averaged 28.8°C (±2.02°C, SD) and 26.0°C (±1.01°C, SD) at night. Relative humidity was maintained at 63.4% (±9.04%, SD) during the day and 69.7% (±1.73%, SD) at night (Rotation Atomiser Defensor ABS3, Condair Ltd., Pfäffikon, Schwyz, Switzerland).

Eight DAT, each pot was watered to a uniform weight so that approximately 600 ml of water was maintained in the soil. For the salt stressed plants, 100 ml of NaCl solution (270 mM NaCl: 9.9 mM $CaCl_ß$) was applied to the square dish, small holes in the bottom of the pots allowed for the infiltration of salt into the soil through capillary action. Salt treatment was applied in two

steps of 45 mM to reach a final concentration of 90 mM, at 10 and 13 DAT (Supplemental Fig. S1A). Control plants received 100 mL of water on day 10 and 13 (Supplemental Fig. S2).

The experiment involved two Smarthouses that were used consecutively for 3 periods, the periods forming blocks. In each Smarthouse, 432 pots were situated in 24 lanes × 18 positions. A split-plot design was employed with two consecutive pots having the same genotype, but with the two different salt treatments randomly assigned to them. For each period, a blocked, partially-replicated design was used to allocate the 378 genotypes to the 432 pairs of pots in the two Smarthouses and was generated using DiGGer, a package for the R statistical computing environment (Coombes, 2009; R Core Team, 2014). There were 5 check genotypes that were always included in a Smarthouse and 373 genotypes of which 49 were duplicated in a period and the remaining 324 were unreplicated. The duplicated genotypes differed between periods so that 147 genotypes were duplicated in total. The entire dataset of RGB and FLUO images can be accessed through the iPlant Collaborative (http://mirrors.iplantcollaborative.org/browse/iplant/home/shared/walia_rice_salt).

5.4.2 VIS/RGB image acquisition and processing

Plant imaging was initiated two days before the first salt application (8 DAT) to provide a baseline for the determination of growth rate. To assess plant growth and morphological traits plants were imaged in an imaging chamber using a 5 mega pixel visible/RGB camera (Basler pilot piA2400-12gc, Basler AG, Ahrensburg, Germany; Supplemental Fig. S1C). Lighting conditions, plant positioning and camera settings were fixed throughout the experiment. Plants were imaged for 14 d after the final salt application (until 27 DAT). For each plant, two side-view images, in which the pot was rotated 90°, and a single top view image were acquired daily.

The 142,671 color (RGB) images were processed in LemnaGrid (LemnaTec GmbH, Aachen, Germany) to remove non-plant pixels from images. RGB processing consisted of two main steps: (1) color classification for object extraction and (2) noise reduction (Supplemental Fig. S21). To extract plant-pixels from the background, the colors of the image were assigned to object (plant) and background (non-plant) color classes (Supplemental Fig. S22). Briefly, a set of predefined colors was selected manually to represent the range of colors present in shoot tissues. For each image, pixels are assigned to color classes using the nearest neighbor (NN) method. The NN method searches in a set of predefined colors in order to find the most similar (nearest neighbor), defined as the smallest Euclidean distance, for any given pixel. Pixels are then assigned membership to the most similar predefined colors. Non-plant pixels were further removed from the processed images using a series of erosion and dilation steps. Seven traits were derived from RGB images and were used to describe plant growth (Table S1). To determine the effects of salinity treatment on each trait, a one-way ANOVA was performed at each individual time-point, where treatment was considered as a fixed effect and accession was considered as a random effect. Raw p-values were adjusted using a Bonferroni correction of $\alpha = 0.05$, which corresponds to a raw p-value < 0.0028.

5.4.3 Fluorescence image acquisition and processing

To assess the effects of salinity stress on chlorophyll content or leaf senescence, plants were imaged in a separate fluorescence-imaging chamber illuminated with a constant blue light (400-500 nm). Fluorescence images were captured using a 1.4 mega pixel camera (Basler scout scA1400-17gc, Basler AG, Ahrensburg, Germany) with a high pass filter, which captures steady state chlorophyll fluorescence from 500 nm to 700 nm. To prevent the detection of fluorescence

signals from algae only images captured from two-side view angles were used for further analysis (Supplemental Fig. S1C).

To process the 95,118 fluorescence images we developed an open source software called “Image Harvest” (https://git.unl.edu/aknecht2/ih). “Image Harvest” is written in Python and utilizes functions in OpenCV to extract plant pixels from LemnaTec images (Bradski,). For the execution of the image processing pipelines on high performance computing clusters “Image Harvest” utilizes Pegasus, which translates a series of computational tasks into a Directed Acyclic Graph and utilizes HTCondor to execute the jobs in parallel on the Open Science Grid (Thain et al., 2005; Sfiligoi et al., 2009; Deelman et al., 2014). The processing workflow was developed using a set of 95,118 side view images, and the quality of each processing step was determined manually. The workflow consists of three main steps: pot masking, background removal and image cropping. The background removal steps are accomplished using the `'colorFilter'` function. Due to the bright plant and dark background, filtering pixels by intensity serves as the primary tool to distinguish between plant and non-plant pixels. Specifically, filtering based on the net intensity (r + g + b), and filtering by the difference between the red and green channels are the major operations. The first filtering keeps pixels from both the plant and pot, whereas the second threshold removes the green pixels of the pot, as well as orange pixels that may show up from debris fluorescing in the image. The image is then cropped based on contours (connected areas of pixels) in the image, through the use of the `'contourCut'` function.

To identify color ranges that may be representative of senescent tissue, histograms representing the range of pixel values for red and green color channels present in all fluorescence images were generated and segmented into 90 color classes. All possible pairwise combinations of R and G

classes were generated, resulting in a total of 90 final color classes that represented the full range of color values in each fluorescence image. For each image, pixels were allocated to each bin depending on the color value. The number of pixels for each bin were summed from both side view images and expressed as a percentage of the plant area from the two side view images. To identify color classes displaying differences between treatments across all 373 accessions, a one-way blocked ANOVA was performed at each individual time-point, where treatment was considered as a fixed effect and accession was as a block. Raw *p*-values were adjusted using a Bonferroni correction of $\alpha = 0.05$, which corresponds to a raw *p*-value < 0.00056.

5.4.4 Determination of sodium and potassium content

The newest expanding leaf (third leaf) was marked on each plant at the time of the first salt application. At the end of the experiment this leaf was harvested, rinsed with milliQ water and patted dry. The samples were placed into 50 ml conical tubes and after being dried at 60ºC overnight, the dry weight was determined for each leaf sample. Leaves were digested in 10 ml of 1% nitric acid (70% Nitric Acid; Chem-Supply NA001-500M, Gillman, SA, AU) at 70ºC for 8 h. Samples were diluted in milliQ water at a 1:5 or 1:10 ratio, and Na^+ and K^+ content were determined against the appropriate Na^+ and K^+ standard (100:1000-500:1000 µM NaCl:KCl) using flame photometry (Model 420 Flame Photometer, Sherwood Scientific Ltd., Cambridge, UK) . Ion content was calculated per g of dry mass as described by Munns et al (Munns, 2010).

5.4.5 Hierarchical clustering analysis of fluorescence color classes

Hierarchical clustering analysis was used to identify and visualize similar temporal trends between fluorescence color classes. For each color class and treatment the mean response across all 373 accessions was determined. Since the values differ by orders of magnitude between color classes the raw mean values were transformed to Z-scores using the `'scale'` function in R (R

Core Team, 2014). Clustering was done with the complete-linkage method using Dynamic Time Warping (DTW) as a distance metric.

5.4.6 Canonical correlation analysis

Pearson correlation analysis was conducted between fluorescence and growth image-based metric at each time point to examine the relationships between traits. At each time point, correlation analysis was done using the `'rcorr'` function with the `'Pearson'` option in the `'Hmisc'` package in R (Jr et al., 2014; R Core Team, 2014).

5.4.7 Comparison of salinity induced growth response models between subpopulations

To determine whether there were any differences in the salinity response among the five subpopulations (*admix*, *aus*, *temperate japonica*, *tropical japonica* and *indica*) the mean salinity-induced growth response, defined here as the square root transformed ratio of the projected shoot area (PSA) of plants under salt stress over that of plants under normal growth situations, was calculated for each subpopulation. *Aromatic* accessions were excluded from the analysis due to low *n*. The mean response for each subpopulation was fitted to a decreasing logistic function with a first order autoregressive covariance structure over time using `'gnls'` and `'SSLogis'` functions in the `'nlme'` package in R (R Core Team, 2014; Pinheiro et al., 2015). For each pairwise comparison between subpopulations two models were fit: one where the parameters of the model are considered to be different between subpopulations, and one where the parameters are considered to be similar between subpopulations. An ANOVA was used to test the null hypothesis that the parameters are similar between subpopulations.

5.4.8 Longitudinal salinity-induced growth responses association analysis

To identify the SNPs associated with characteristics of the growth pattern, a novel functional association-mapping model was applied (Supplemental Fig. S1C). The salt stress induced growth response is measured as the square root transformed ratio of the projected shoot area (PSA) of plants under salt stress over that of plants under normal growth situations. For each line and a specific SNP, the growth response over 18 days is fitted with a decreasing logistic curve for each genotype, with a first order autoregressive covariance structure over time. For each SNP, a likelihood ratio test is carried out on whether the growth response curves are the same for the two genotypes. At each time point, on the other hand, the relatedness between lines is accounted for by a random line effect with the covariance matrix proportional to the kinship matrix estimated by the software package EMMA (Kang et al., 2008).

In summary, for each line, the growth response is modeled with a nonlinear mixed model. While at each time point, the model is the linear mixed model commonly used for plant association studies as in Kang et al. and others using the kinship matrix between lines to account for the population structure (Yu et al., 2006; Malosetti et al., 2007; Zhao et al., 2007, 2011; Kang et al., 2008; Huang et al., 2010; Famoso et al., 2011).. For each line across time points, on the other hand, the model reduces to a nonlinear mixed model with a first order autoregressive covariance structure. Taken together, this results in an extended mixed model with correlation structure both across time and also across lines. This novel approach is required to account for the two types of dependency structure in order to obtain valid inference results. Further details on the extended mixed model can be found in the Supplemental Methods.

From the 44,000 SNPs on the array, we removed SNPs with missing values for more than 10% of the lines and those with minor allele frequency less than 10% (Zhao et al., 2011). This results

in 26,258 SNPs for association analysis. For each SNP, a likelihood ratio test is carried out on whether the growth response curves are the same for the two genotypes. The p-values from the analysis of three replicates are combined with Fisher's method to obtain the combined *p*-value, which is subsequently used to identify significant SNPs (Hartung et al., 2011). Significant SNPs within a 200 Kb window were combined and considered as a single QTL. The 200 Kb window was chosen based on the estimated LD decay in this diversity panel (Zhao et al., 2011).

5.4.9 Conventional mixed model genome-wide association analysis

A conventional mixed model genome-wide association analysis was used to identify genomic regions associated with leaf ion content (Na^+, Na^+:K^+ and K^+) and fluorescence imaging traits. The mixed linear model can be summarized as: $y = \mathbf{X\beta} + C\boldsymbol{\gamma} + \boldsymbol{Z}u + \boldsymbol{e}$. Where β and γ represent coefficient vectors for SNP effects and subpopulation principal components respectively, which are fixed effects. u is a random effect that accounts for population structure and relatedness. While *Z* represents the corresponding design matrices, and *e* is the random error term. The mixed model was implemented using EMMA in R using the same 26,258 SNPs that were used for longitudinal salinity-induced growth response association analysis (Kang et al., 2008).

5.5 References

Apse, M.P., G.S. Aharon, W. a Snedden, and E. Blumwald. 1999. Salt Tolerance Conferred by Overexpression of a Vacuolar Na+/H+ Antiport in Arabidopsis. Science 285(5431): 1256–1258Available at http://www.sciencemag.org/cgi/doi/10.1126/science.285.5431.1256.

Asano, T., N. Hayashi, M. Kobayashi, N. Aoki, A. Miyao, I. Mitsuhara, H. Ichikawa, S. Komatsu, H. Hirochika, S. Kikuchi, and R. Ohsugi. 2012. A rice calcium-dependent protein kinase OsCPK12 oppositely modulates salt-stress tolerance and blast disease resistance. Plant J. 69(1): 26–36.

Azaizeh, H., and E. Steudle. 1991. Effects of Salinity on Water Transport of Excised Maize (Zea mays L .) Roots ' Federal Republic of Germany. : 1136–1145.

Bac-Molenaar, J. a., D. Vreugdenhil, C. Granier, and J.J.B. Keurentjes. 2015. Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. J. Exp. Bot.

Berger, B., B. Parent, and M. Tester. 2010. High-throughput shoot imaging to study drought responses. J. Exp. Bot. 61(13): 3519–28.

Berthomieu, P., G. Conéjéro, A. Nublat, W.J. Brackenbury, C. Lambert, C. Savio, N. Uozumi, S. Oiki, K. Yamada, F. Cellier, F. Gosti, T. Simonneau, P.A. Essah, M. Tester, A.A. Véry, H. Sentenac, and F. Casse. 2003. Functional analysis of AtHKT1 in Arabidopsis shows that Na+ recirculation by the phloem is crucial for salt tolerance. EMBO J. 22(9): 2004–2014.

Bonilla, P., D. Mackell, K. Deal, and G. Gregorio. 2002. RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (Oryza sativa L.) using recombinant inbred lines. Philipp. Agric. Sci. 85: 68–76.

Boursiac, Y., J. Boudet, O. Postaire, D.-T. Luu, C. Tournaire-Roux, and C. Maurel. 2008. Stimulus-induced downregulation of root water transport involves reactive oxygen species-activated cell signalling and plasma membrane intrinsic protein internalization. Plant J. 56(2): 207–218.

Boursiac, Y., S. Chen, D.-T. Luu, M. Sorieul, N. Van Den Dries, C. Maurel, N. Van Den Dries, and C. Maurel. 2005. Early Effects of Salinity on Water Transport in Arabidopsis Roots . Molecular and Cellular Features of Aquaporin Expression 1. Plant Physiol. 139(October): 790–805Available at http://www.ncbi.nlm.nih.gov/pubmed/16183846.

Bradski, G. OpenCV. Dr. Dobb's J. Softw. Tools.

Brown, T.B., R. Cheng, X.R.R. Sirault, T. Rungrat, K.D. Murray, M. Trtilek, R.T. Furbank, M. Badger, B.J. Pogson, and J.O. Borevitz. 2014. TraitCapture: Genomic and environment modelling of plant phenomic data. Curr. Opin. Plant Biol. 18(1): 73–79Available at http://dx.doi.org/10.1016/j.pbi.2014.02.002.

Busemeyer, L., A. Ruckelshausen, K. Möller, A.E. Melchinger, K. V Alheit, H.P. Maurer, V. Hahn, E. a Weissmann, J.C. Reif, and T. Würschum. 2013. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. Sci. Rep.

3: 2442Available at http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3743059&tool=pmcentrez&rendertype=abstract.

Campo, S., P. Baldrich, J. Messeguer, E. Lalanne, M. Coca, and B. San Segundo. 2014. Overexpression of a Calcium-Dependent Protein Kinase Confers Salt and Drought Tolerance in Rice by Preventing Membrane Lipid Peroxidation. Plant Physiol. 165(2): 688–704Available at http://www.ncbi.nlm.nih.gov/pubmed/24784760.

Carvajal, M., V. Martínez, and C.F. Alcaraz. 1999. Physiological function of water channels as affected by salinity in roots of paprika pepper. Physiol. Plant. 105(1): 95–101.

Cheeseman, J.M. 1988. Mechanisms of salinity tolerance in plants. Plant Physiol. 87(3): 547–50Available at http://www.ncbi.nlm.nih.gov/pubmed/16666181%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1054794.

Chen, D., K. Neumann, and S. Friedel. 2014. Dissecting the Phenotypic Components of Crop Plant Growth and Drought Responses Based on High-Throughput Image Analysis. Plant Cell … 26(12): 4636–4655.

Choi, W.-G., M. Toyota, S.-H. Kim, R. Hilleary, and S. Gilroy. 2014. Salt stress-induced Ca2+ waves are associated with rapid, long-distance root-to-shoot signaling in plants. Proc. Natl. Acad. Sci. U. S. A. 111(17): 6497–502.

Coombes, N. 2009. Digger design search tool in R.

Cui, Y., J. Zhu, and R. Wu. 2006. Functional mapping for genetic control of programmed cell death. (5): 458–469.

Das, K., J. Li, Z. Wang, C. Tong, G. Fu, Y. Li, M. Xu, K. Ahn, D. Mauger, R. Li, and R. Wu. 2011. A dynamic model for genome-wide association studies. Hum. Genet. 129(6): 629–639.

Davenport, R.J., A. Muñoz-Mayor, D. Jha, P.A. Essah, A. Rus, and M. Tester. 2007. The Na+ transporter AtHKT1;1 controls retrieval of Na + from the xylem in Arabidopsis. Plant, Cell Environ. 30(2007): 497–507.

Deelman, E., K. Vahi, G. Juve, M. Rynge, S. Callaghan, P.J. Maechling, R. Mayani, W. Chen, R.F. da Silva, M. Livny, and others. 2014. Pegasus, a workflow management system for science automation. Futur. Gener. Comput. Syst.

Famoso, A.N., K. Zhao, R.T. Clark, C.-W. Tung, M.H. Wright, C. Bustamante, L. V Kochian, and S.R. McCouch. 2011. Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7(8): e1002221.

Fricke, W. 2004. Rapid and tissue-specific accumulation of solutes in the growth zone of barley leaves in response to salinity. Planta 219(3): 515–25Available at http://www.ncbi.nlm.nih.gov/pubmed/15085434.

Fricke, W., and W.S. Peters. 2002. The biophysics of leaf growth in salt-stressed barley. A study at the cell level. Plant Physiol. 129(May): 374–388.

Ghanem, M.E., A. Albacete, C. Martínez-Andújar, M. Acosta, R. Romero-Aranda, I.C. Dodd, S. Lutts, and F. Pérez-Alfocea. 2008. Hormonal changes during salinity-induced leaf senescence in tomato (Solanum lycopersicum L.). J. Exp. Bot. 59(11): 3039–3050.

Gierth, M., and P. Mäser. 2007. Potassium transporters in plants--involvement in K+ acquisition, redistribution and homeostasis. FEBS Lett. 581(12): 2348–56.

Golzarian, M.R., R. a Frick, K. Rajendran, B. Berger, S. Roy, M. Tester, and D.S. Lun. 2011. Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods 7(1): 2.

Halfter, U., M. Ishitani, and J.K. Zhu. 2000. The Arabidopsis SOS2 protein kinase physically interacts with and is activated by the calcium-binding protein SOS3. Proc Natl Acad Sci 97(7): 3735–3740Available at http://dx.doi.org/10.1073/pnas.040577697.

Hartung, J., G. Knapp, and B.K. Sinha. 2011. Statistical meta-analysis with applications. John Wiley & Sons.

He, Q., A. Berg, Y. Li, C.E. Vallejos, and R. Wu. 2010. Mapping genes for plant structure, development and evolution: functional mapping meets ontology. Trends Genet. 26(1): 39–46.

Honsdorf, N., T. March, B. Berger, M. Tester, and K. Pillen. 2014. High-Throughput Phenotypig to Detect Drought Tolerance QTL in Barley Introgression Lines. PLoS One 9(5): e97047.

Horie, T., T. Kaneko, G. Sugimoto, S. Sasano, S.K. Panda, M. Shibasaka, and M. Katsuhara. 2011. Mechanisms of water transport mediated by PIP aquaporins and their regulation via phosphorylation events under salinity stress in barley roots. Plant Cell Physiol. 52(4): 663–675.

Horie, T., I. Karahara, and M. Katsuhara. 2012. Salinity tolerance mechanisms in glycophytes: An overview with the central focus on rice plants. Rice 5(1): 11Available at ???

Huang, S., and W. Spielmeyer. 2006. A sodium transporter (HKT7) is a candidate for Nax1, a gene for salt tolerance in durum wheat. Plant Phys 142(4): 1718–27.

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, M. Li, D. Fan, Y. Guo, A. Wang, L. Wang, L. Deng, W. Li, Y. Lu, Q. Weng, K. Liu, T. Huang, T. Zhou, Y. Jing, W. Li, Z. Lin, E.S. Buckler, Q. Qian, Q. Zhang, J. Li, and B. Han. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Publ. Gr. 42(11): 961–967Available at http://dx.doi.org/10.1038/ng.695.

Ishitani, M., J. Liu, U. Halfter, C.S. Kim, W. Shi, and J.K. Zhu. 2000. SOS3 function in plant salt tolerance requires N-myristoylation and calcium binding. Plant Cell 12(September): 1667–1678.

Jr, F.E.H., with contributions from Charles Dupont, and many others. 2014. Hmisc: Harrell Miscellaneous.

Kang, H.M., N. a Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178(3): 1709–23.

Koyama, M.L., A. Levesley, R.M.D. Koebner, T.J. Flowers, and A.R. Yeo. 2001. Quantitative Trait Loci for Component Physiological Traits Determining Salt Tolerance in Rice 1. 125(January): 406–422.

Latz, A., N. Mehlmer, S. Zapf, T.D. Mueller, B. Wurzinger, B. Pfister, E. Csaszar, R. Hedrich, M. Teige, and D. Becker. 2013. Salt stress triggers phosphorylation of the Arabidopsis vacuolar K+ channel TPK1 by calcium-dependent protein kinases (CDPKs). Mol. Plant 6(4): 1274–89.

Lenth, R. V, and M. Hervé. 2015. lsmeans: Least-Squares Means.

Lin, H.X., M.Z. Zhu, M. Yano, J.P. Gao, Z.W. Liang, W. a Su, X.H. Hu, Z.H. Ren, and D.Y. Chao. 2004. QTLs for Na+ and K+ uptake of the shoots and roots controlling rice salt tolerance. Theor. Appl. Genet. 108(2): 253–60.

Liu, C., B. Mao, S. Ou, W. Wang, L. Liu, Y. Wu, C. Chu, and X. Wang. 2014. OsbZIP71, a bZIP transcription factor, confers salinity and drought tolerance in rice. Plant Mol. Biol. 84(1–2): 19–36.

Lutts, S., J.M. Kinet, and J. Bouharmont. 1996. NaCl-induced Senescence in Leaves of Rice (Oryza sativa L.) Cultivars Differing in Salinity Resistance. Ann. Bot. 78(3): 389–398Available at http://aob.oxfordjournals.org/content/78/3/389.abstract.

Malosetti, M., C.G. van der Linden, B. Vosman, and F.A. van Eeuwijk. 2007. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to Phytophthora infestans in potato. Genetics 175(2): 879–889.

Martínez-Ballesta, M.C., F. Aparicio, V. Pallás, V. Martínez, and M. Carvajal. 2003. Influence of saline stress on root hydraulic conductance and PIP expression inArabidopsis. J. Plant Physiol. 160(6): 689–697.

Matsuda, K., and a Riazi. 1981. Stress-induced osmotic adjustment in growing regions of barley leaves. Plant Physiol. 68: 571–576.

Mian, A., R.J.F.J. Oomen, S. Isayenkov, H. Sentenac, F.J.M. Maathuis, and A.-A. Véry. 2011. Over-expression of an Na+-and K+-permeable HKT transporter in barley improves salt tolerance. Plant J. 68(3): 468–479.

Møller, I.S., M. Gilliham, D. Jha, G.M. Mayo, S.J. Roy, J.C. Coates, J. Haseloff, and M. Tester. 2009. Shoot Na+ exclusion and increased salinity tolerance engineered by cell type-specific alteration of Na+ transport in Arabidopsis. Plant Cell 21(July): 2163–2178.

Moore, C.R., L.S. Johnson, I.-Y. Kwak, M. Livny, K.W. Broman, and E.P. Spalding. 2013. High-throughput computer vision introduces the time axis to a quantitative trait map of a plant growth response. Genetics 195(3): 1077–86.

Munns, R. 2002. Comparative physiology of salt and water stress. Plant. Cell Environ. 25(2):

239–250Available at http://www.ncbi.nlm.nih.gov/pubmed/11841667.

Munns, R. 2010. Plant Stress Tolerance (R Sunkar, Ed.). 639(1): 25–38.

Munns, R., H. Greenway, R. Delane, and J. GIBBS. 1982. Ion Concentration and Carbohydrate Status of the Elongating Leaf Tissue of Hordeum vulgare Growing at High External NaCl II. Cause of Grwoth Reduction. J. Exp. Bot. 33(4): 574–583.

Munns, R., and R. a. James. 2003. Screening Methods for SalinityTolerance: A Case Study with Tetraploid Wheat. Plant Soil 253(1): 201–218Available at http://link.springer.com/10.1023/A:1024553303144.

Munns, R., R. a James, B. Xu, A. Athman, S.J. Conn, C. Jordans, C.S. Byrt, R. a Hare, S.D. Tyerman, M. Tester, D. Plett, and M. Gilliham. 2012. Wheat grain yield on saline soils is improved by an ancestral Na^+ transporter gene. Nat. Biotechnol. 30(4): 360–4.

Munns, R., and J. Passioura. 1984. Effect of Prolonged Exposure to Nacl on the Osmotic Pressure of Leaf Xylem Sap From Intact, Transpiring Barley Plants. Aust. J. Plant Physiol. 11(1968): 497.

Munns, R., and a Termaat. 1986. Whole-Plant Responses to Salinity. Aust. J. Plant Physiol. 13(1977): 143.

Munns, R., and M. Tester. 2008. Mechanisms of salinity tolerance. Annu. Rev. Plant Biol. 59: 651–81.

Negrão, S., B. Courtois, N. Ahmadi, I. Abreu, N. Saibo, and M.M. Oliveira. 2011. Recent updates on salinity stress in rice: from physiological to molecular responses. CRC. Crit. Rev. Plant Sci. 30(4): 329–377.

Obata, T., H.K. Kitamoto, A. Nakamura, A. Fukuda, and Y. Tanaka. 2007. Rice shaker potassium channel OsKAT1 confers tolerance to salinity stress on yeast and rice cells. Plant Physiol. 144(4): 1978–1985.

Paine, C.E.T., T.R. Marthews, D.R. Vogt, D. Purves, M. Rees, A. Hector, and L. a. Turnbull. 2012. How to fit nonlinear plant growth models and calculate growth rates: An update for ecologists. Methods Ecol. Evol. 3: 245–256.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team. 2015. {nlme}: Linear and Nonlinear Mixed Effects Models.

Platten, J.D., J. a Egdane, and A.M. Ismail. 2013. Salinity tolerance, Na+ exclusion and allele mining of HKT1;5 in Oryza sativa and O. glaberrima: many sources, many genes, one mechanism? BMC Plant Biol. 13(1): 32.

Poiré, R., V. Chochois, X.R.R. Sirault, J.P. Vogel, M. Watt, and R.T. Furbank. 2014. Digital imaging approaches for phenotyping whole plant nitrogen and phosphorus response in Brachypodium distachyon. J. Integr. Plant Biol. 56(8): 781–796.

R Core Team. 2014. R: A Language and Environment for Statistical Computing.

Rajendran, K., M. Tester, and S.J. Roy. 2009. Quantifying the three main components of salinity

tolerance in cereals. Plant, Cell Environ. 32(3): 237–249.

Rawson, H.M., and R. Munns. 1984. Leaf expansion in sunflower as influenced by salinity and short-term changes in carbon fixation. Plant. Cell Environ. 7(3): 207–213.

Ren, Z.-H., J.-P. Gao, L.-G. Li, X.-L. Cai, W. Huang, D.-Y. Chao, M.-Z. Zhu, Z.-Y. Wang, S. Luan, and H.-X. Lin. 2005. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat. Genet. 37(10): 1141–6.

Rivandi, J., J. Miyazaki, M. Hrmova, M. Pallotta, M. Tester, and N.C. Collins. 2011. A SOS3 homologue maps to HvNax4, a barley locus controlling an environmentally sensitive Na + exclusion trait. J. Exp. Bot. 62(3): 1201–1216.

Rus, A., S. Yokoi, A. Sharkhuu, M. Reddy, B.H. Lee, T.K. Matsumoto, H. Koiwa, J.K. Zhu, R. a Bressan, and P.M. Hasegawa. 2001. AtHKT1 is a salt tolerance determinant that controls Na(+) entry into plant roots. Proc. Natl. Acad. Sci. U. S. A. 98: 14150–14155.

Schmidt, R., D. Mieulet, H.-M. Hubberten, T. Obata, R. Hoefgen, A.R. Fernie, J. Fisahn, B. San Segundo, E. Guiderdoni, J.H.M. Schippers, and B. Mueller-Roeber. 2013. SALT-RESPONSIVE ERF1 Regulates Reactive Oxygen Species – Dependent Signaling during the Initial Response to Salt Stress in Rice. Plant Cell 25(June): 2115–2131.

Sfiligoi, I., D.C. Bradley, B. Holzman, P. Mhashilkar, S. Padhi, and F. Würthwein. 2009. The pilot way to Grid resources using glideinWMS. 2009 WRI World Congr. Comput. Sci. Inf. Eng. CSIE 2009 2: 428–432.

Shi, H., M. Ishitani, C. Kim, and J.K. Zhu. 2000. The Arabidopsis thaliana salt tolerance gene SOS1 encodes a putative Na+/H+ antiporter. Proc. Natl. Acad. Sci. U. S. A. 97(12): 6896–6901.

Slovak, R., C. Göschl, X. Su, K. Shimotani, T. Shiina, and W. Busch. 2014. A scalable open-source pipeline for large-scale root phenotyping of Arabidopsis. Plant Cell Online 26(6): 2390–2403.

Sunarpi, T. Horie, J. Motoda, M. Kubo, H. Yang, K. Yoda, R. Horie, W.-Y. Chan, H.-Y. Leung, K. Hattori, M. Konomi, M. Osumi, M. Yamagami, J.I. Schroeder, and N. Uozumi. 2005. Enhanced salt tolerance mediated by AtHKT1 transporter-induced Na unloading from xylem vessels to xylem parenchyma cells. Plant J. 44(6): 928–38.

Székely, G., E. Ábrahám, Á. Cséplo, G. Rigó, L. Zsigmond, J. Csiszár, F. Ayaydin, N. Strizhov, J. Jásik, E. Schmelzer, C. Koncz, and L. Szabados. 2008. Duplicated P5CS genes of Arabidopsis play distinct roles in stress regulation and developmental control of proline biosynthesis. Plant J. 53(1): 11–28.

Thain, D., T. Tannenbaum, and M. Livny. 2005. Distributed computing in practice: The Condor experience. Concurr. Comput. Pract. Exp. 17(2–4): 323–356.

Thomson, M.J., M. Ocampo, J. Egdane, M.A. Rahman, A.G. Sajise, D.L. Adorada, E. Tumimbang-Raiz, E. Blumwald, Z.I. Seraj, R.K. Singh, G.B. Gregorio, and A.M. Ismail. 2010. Characterizing the Saltol Quantitative Trait Locus for Salinity Tolerance in Rice. Rice 3(2–3): 148–160.

Topp, C.N., A.S. Iyer-Pascuzzi, J.T. Anderson, C.-R. Lee, P.R. Zurek, O. Symonova, Y. Zheng, A. Bucksch, Y. Mileyko, T. Galkovskyi, B.T. Moore, J. Harer, H. Edelsbrunner, T. Mitchell-Olds, J.S. Weitz, and P.N. Benfey. 2013. 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. Proc. Natl. Acad. Sci. U. S. A. 110(18): E1695-704.

Walter, A., H. Spies, S. Terjung, R. Kusters, N. Kirchgessner, and U. Schurr. 2002. Spatio-temporal dynamics of expansion growth in roots: automatic quantification of diurnal course and temperature response by digital image sequence processing. J. Exp. Bot. 53(369): 689–698.

Wang, Z., Z. Chen, J. Cheng, Y. Lai, J. Wang, Y. Bao, J. Huang, and H. Zhang. 2012. QTL Analysis of Na+ and K+ Concentrations in Roots and Shoots under Different Levels of NaCl Stress in Rice (Oryza sativa L.). PLoS One 7(12): 1–9.

van der Weele, C.M., H.S. Jiang, K.K. Palaniappan, V.B. Ivanov, K. Palaniappan, and T.I. Baskin. 2003. A new algorithm for computational image analysis of deformable motion at high spatial and temporal resolution applied to root growth. Roughly uniform elongation in the meristem and also, after an abrupt acceleration, in the elongation zone. Plant Physiol. 132(3): 1138–1148.

Wu, R., and M. Lin. 2006. Functional mapping—how to map and study the genetic architecture of dynamic complex traits. Nat. Rev. Genet. 7(March): 229–237.

Würschum, T., W. Liu, L. Busemeyer, M.R. Tucker, J.C. Reif, E. a Weissmann, V. Hahn, A. Ruckelshausen, and H. Maurer. 2014. Mapping dynamic QTL for plant height in triticale. BMC Genet. 15: 59.

Xu, Z., X. Shen, and W. Pan. 2014. Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. PLoS One 9(8): e102312.

Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang, W. Fang, H. Feng, W. Xie, X. Lian, G. Wang, Q. Luo, Q. Zhang, Q. Liu, and L. Xiong. 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat. Commun. 5: 5087.

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, and others. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38(2): 203–208.

Zeng, L., and M. Shannon. 2000. Salinity effects on seedling growth and yield components of rice. Crop Sci. 40: 996–1003Available at https://dl.sciencesocieties.org/publications/cs/abstracts/40/4/996.

Zeng, L., M.C. Shannon, and C.M. Grieve. 2002. Evaluation of salt tolerance in rice genotypes by multiple agronomic parameters. : 235–245.

Zhang, X., R.J. Hause, and J.O. Borevitz. 2012. Natural genetic variation for growth and

development revealed by high-throughput phenotyping in Arabidopsis thaliana. G3 Genes| Genomes| Genet. 2(1): 29–34.

Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, P. Marjoram, and others. 2007. An Arabidopsis example of association mapping in structured samples. PLoS Genet. 3(1): e4.

Zhao, K., C.-W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2(1–4): 467.

5.6 Figures

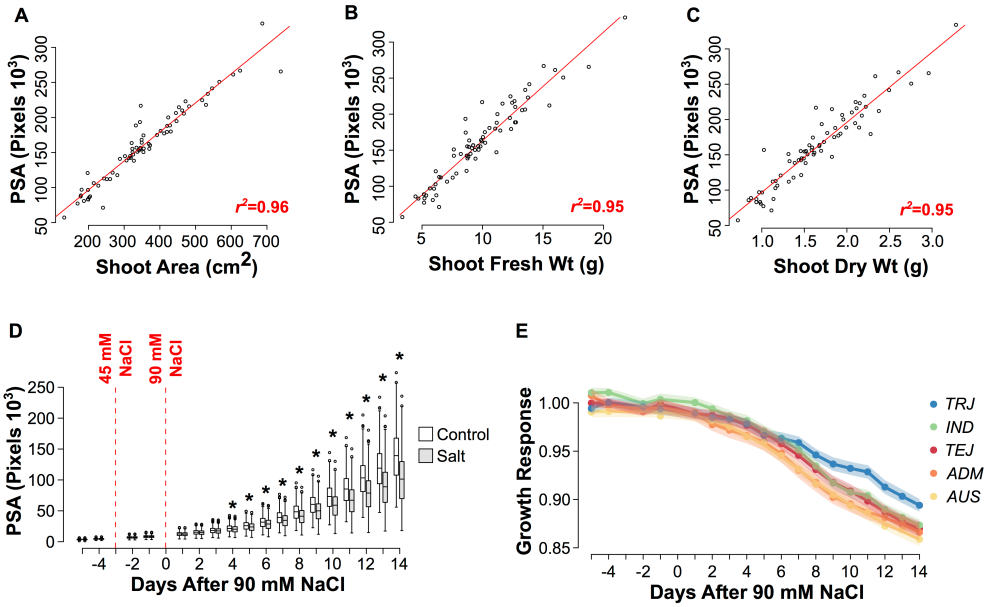


Figure 1. Salinity-induced growth responses in rice diversity panel. (A-C) Relationship between projected shoot area (PSA) and conventional biomass metrics. Pearson correlation analyses were performed between PSA and shoot area (A), shoot fresh mass (B) and shoot dry mass (C). Comparisons of PSA between treatments at each of the 18 days of imaging (D). Differences between treatments at each time point and was determined using a one-way blocked ANOVA where accession is considered as a block ($p < 0.0027$). (E) Comparison of salinity-induced growth response models between each of the five subpopulations defined by Zhao *et al*

(Zhao et al., 2011). The salinity induced growth response was modeled with a decreasing logistic curve and pair-wise comparisons were made between each subpopulation. Aromatic accessions were excluded due to low *n*. Mean growth responses for each subpopulation are denoted by solid lines, while the standard error for each subpopulation is indicated by shadows. *TRJ: Tropical Japonica; TEJ: Temperate Japonica; IND: Indica; ADM: Admix*

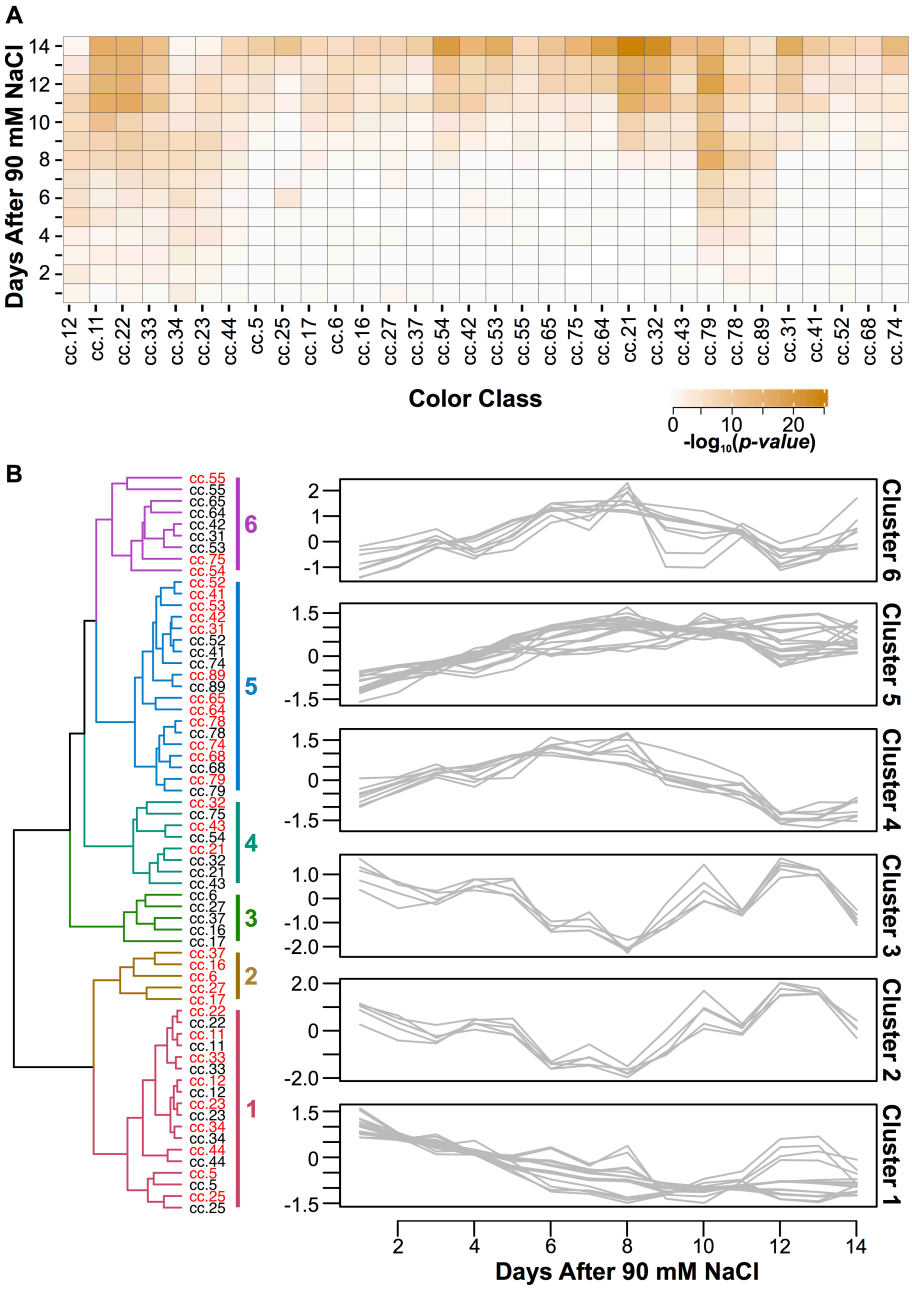


Figure 2. The development of image-based fluorescence traits for monitoring chlorophyll responses to salinity. (A) Salinity-responsive color classes were identified through comparisons between treatments at each time point via one-way ANOVA. Color classes were considered to be responsive to saline conditions if significant differences between treatments were observed in three or more days of 90 mM NaCl stress ($p < 0.00056$). (B) Identification of color classes

exhibiting similar trends over 14 days of 90 mM NaCl. HCA with complete linkage was performed using the mean value in each treatment for each color class. The six clusters are depicted to the right of the dendrogram. Labels in red indicate mean response in saline conditions, while those in black indicate control conditions. The right panel illustrates summarizes the temporal trend captured by each cluster. The mean values for each color class were scaled and centered prior to clustering, so that the mean is zero and variance is one, and are represented on the y-axis.

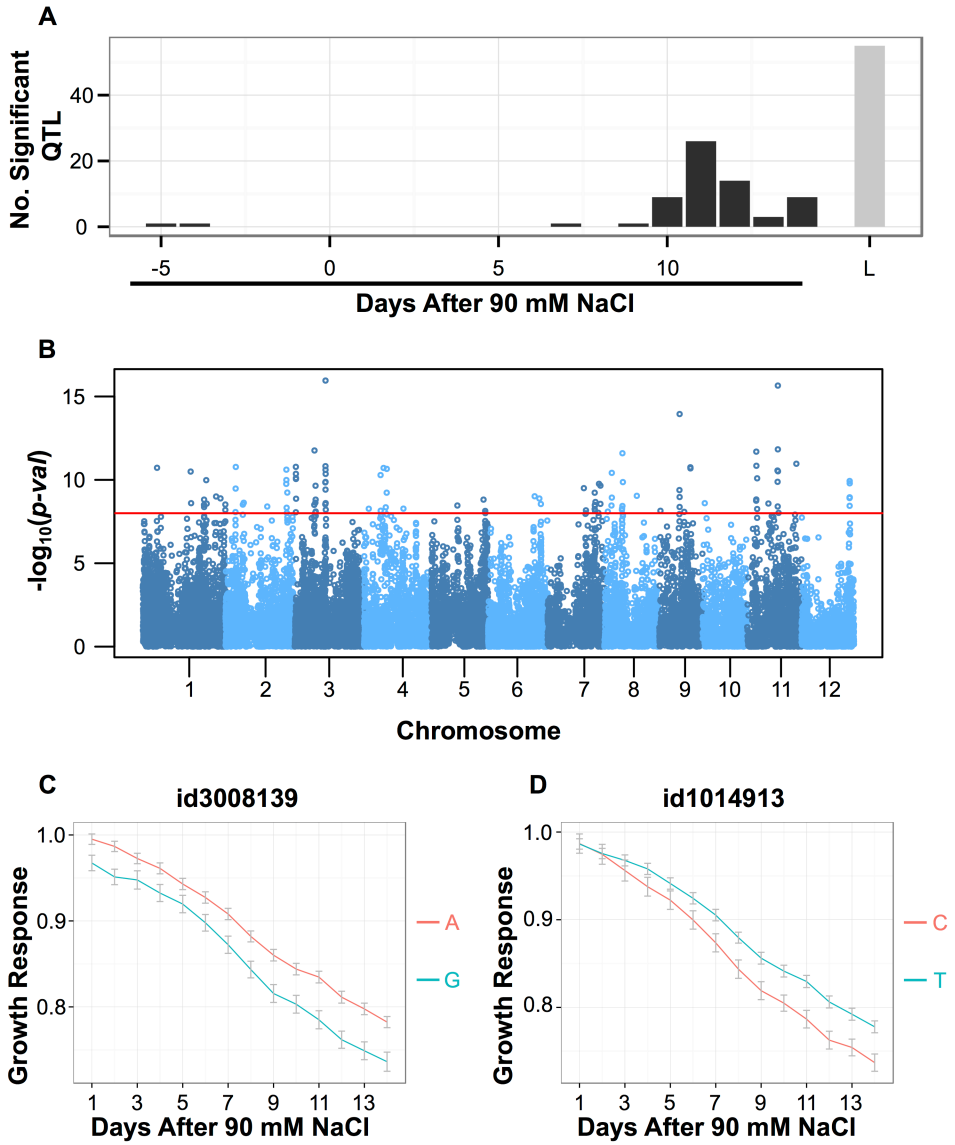


Figure 3. Examining the genetic architecture of salinity-induced growth responses using conventional mixed-model and logistic growth-response association analysis. (A) Comparison of conventional mixed-model association mapping approach with logistic-growth response association mapping approach. A conventional association mapping approach was performed at each time point using the salinity-induced growth response as a phenotypic measure. With the mixed-model approach a SNP was determined to be significant if $p < 10^{-4}$

while a threshold of $p < 10^{-8}$ was used for the logistic growth response model. Significant SNPs within a 200 Kb window were combined and considered as a single QTL. L: Logistic growth-response model (B) Manhattan plot for the logistic growth-response association analysis. The red horizontal line indicates a significance threshold of $p < 10^{-8}$. (C) Comparison of growth response trajectories between allelic groups for the significant association observed at ~16.3 Mb on chromosome 3. A: Major allele (Freq: 0.68), G: Minor allele (Freq: 0.32) (D) Growth trajectories of major and minor allele accessions for the signal observed at ~25 Mb on chromosome 1. T: Major allele (Freq: 0.79), C: Minor allele (Freq: 0.21)

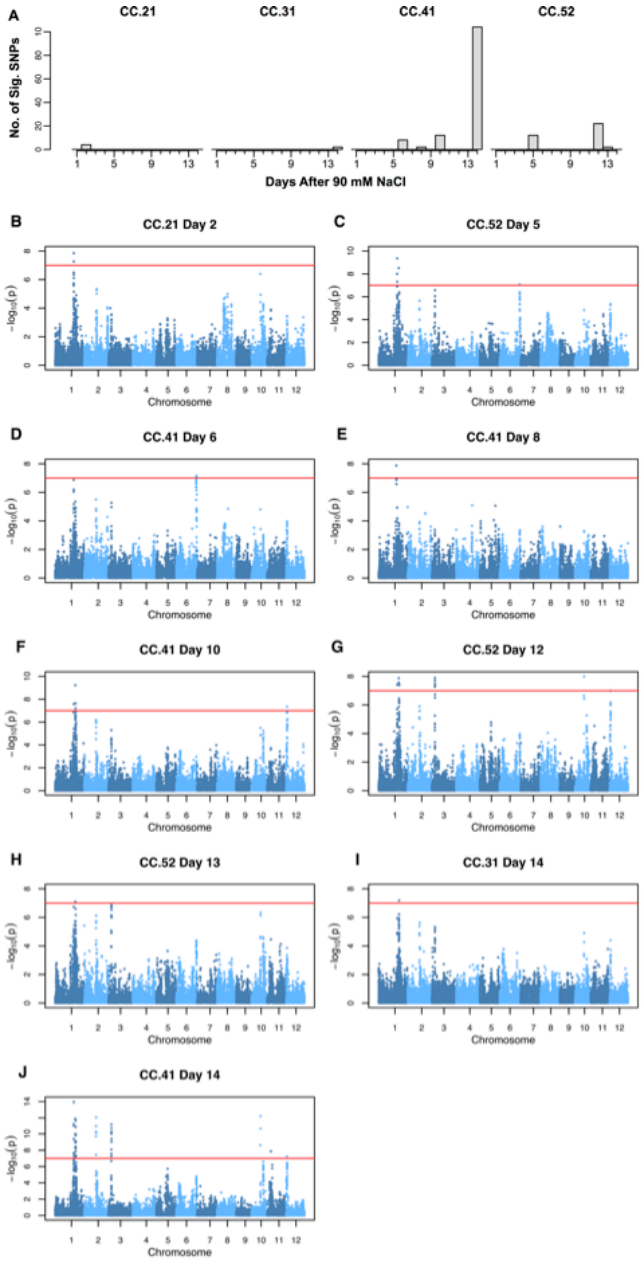


Figure 4. Examining the genetic architecture of salinity-induced fluorescence responses. (A) Summary of the significant signals detected for color classes 21, 31, 41 and 52 at each time point after 90 mM NaCl. The *x*-axis indicates the number of days after 90 mM NaCl application while the y-axis showed the number of SNPs with $p < 10^{-7}$ that were detected with the conventional mixed model GWA approach. (B-J) Genome-wide *p*-values for each fluorescent color class

exhibiting significant genetic associations. The red horizontal line indicates a significance threshold of $p < 10^{-7}$. The corresponding color class and day after 90 mM NaCl is given in the title above each plot.

5.7 Tables

Table I: Comparison of salinity induced growth response models between subpopulations. The mean growth response was calculated for each subpopulation and fitted to a decreasing logistic curve. Aromatic accessions were omitted due to low n. Subpopulations with the same letter indicate no significant differences ($p < 10^{-4}$). Asym: Horizontal asymptote; xmid: value of x at the inflection point of the curve; scal: rate coefficient. TeJ: temperate japonica; TrJ: tropical japonica

	Asym	xmid	scal	AIC	BIC	logLik	$p > 10^{-4}$
Admix	0.997	32.4	-6.66	-3560.161	-3536.05	1785.08	A
Aus	0.999	28.9	-4.82	-4015.505	-3991.30	2012.75	A, B
Indica	0.979	30.0	-5.14	-5152.778	-5126.87	2581.39	A
TeJ	0.982	28.7	-4.21	-6093.456	-6066.51	3051.73	B
TrJ	0.983	30.8	-4.49	-5465.48	-5439.12	2737.74	

5.8 Supplemental Materials

5.8.1 Supplemental Methods: Extended mixed model for association mapping with growth response curve

Recall that the salt stress induced growth response is measured as the square root transformed ratio of the projected shoot area (PSA) of plants under salt stress over that of plants under normal growth situations. For each single SNP, the mean response curve of a specific genotype *k* (1or 2 for the two alleles) is modeled as a decreasing logistic function:

$$f_k(t) = \gamma + \frac{(1+\beta_k)(1-\gamma_k)}{1+\beta_k e^{\alpha_k t}}$$

where *t* is the time, α, β and γ are regression coefficients. Note that this function is nonincreasing, it is 1 at time zero, and it will flatten out as time becomes really long. These are reasonable assumptions for the salt response curve as all plants are under normal growing conditions at time zero, the salt stress has a detrimental effect on the growth on average, and growth will slow for all plants near the end of the growing period.

The growth response for line *i* with genotype *k* at time *t* (day 0 to day 17) is then designated as

$$y_{it} = f_k(t) + \varepsilon_{it}.$$

Here ε_{it} is the error term that incorporates the correlation structure both through time and between lines (kinship). Let $\varepsilon_i = (\varepsilon_{i0}, ..., \varepsilon_{i17})$ be the vector of responses for line *i*, then

$$Cor(\varepsilon_i) = \begin{bmatrix} 1 & \rho & ... & \rho^{17} \\ \rho & 1 & ... & \rho^{16} \\ ... & ... & ... & ... \\ \rho^{16} & \rho^{15} & ... & 1 \end{bmatrix}.$$

Thus, for a single line, the model reduces to a nonlinear mixed model with a first order auto regressive covariance structure. On the other hand, let $\varepsilon_t = (\varepsilon_{1t},...,\varepsilon_{360t})$ be the vector of the response of all lines at time *t*, we have

$$Cov(\varepsilon_t) = \sigma_g^2 K + \sigma_e^2 I$$

where *K* is the kinship matrix between lines, *I* is a 360 by 360 identity matrix, σ_g^2 is the genotype variance component, and σ_e^2 is the error variance component. Thus, considering a single time point, we have exactly the same model used by other researchers in plant association mapping studies.

To consider correlation both across time and across lines, let

$$\varepsilon = vec([\varepsilon_{it}])$$

where the vec operator stacks columns of a matrix into a single vector. We assume that the covariance matrix

$$\Sigma = Cov(\varepsilon) = Cov(\varepsilon_t) \otimes Cor(\varepsilon_i)$$

with $\otimes$ being the kronecker product. Further, we assume that the response y_{it} has a normal distribution with the mean defined by the function $f_k(t)$ and the variance matrix given by Σ. As a result, the likelihood ratio test on whether $f_k(t)$ is the same for *k*=1,2 can be carried out as usual with the likelihood defined by the normal distribution.

5.8.2 Statistical analysis of phenotypic data

Data derived from fluorescence and RGB/VIS images were combined across experiments, and a linear model was fit to calculate the adjusted means for each individual accession using the `'lsmeans'` function in the LSMeans package in R (R Core Team, 2014; Lenth and Hervé, 2015). The linear model included experiment and treatment as fixed effects, and accession as a random effect. The adjusted means were used for phenotypic analysis and GWA of fluorescence traits, as well as phenotypic analysis of PSA.

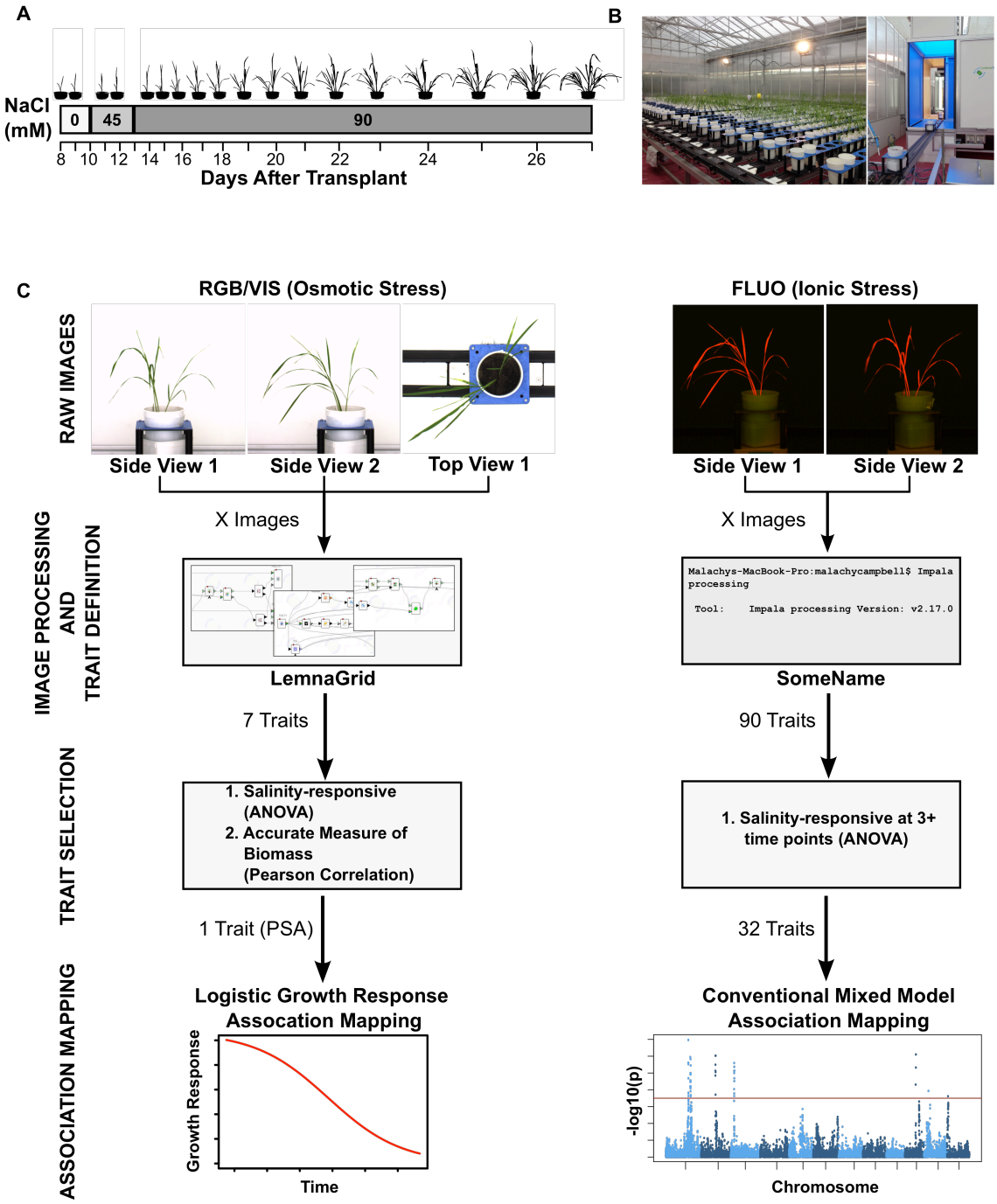


Fig S1: Schematic illustrating the experimental design and pipeline for the analysis of high-throughput phenotyping data. (A) Implementation of 90 mM NaCl. (B) High-throughput phenotyping of rice plants at the early-tillering stage using the LemnaTec system at the Plant

Accelerator. (C) Workflow for image acquisition, processing and analysis (see Materials and Methods for a complete description).

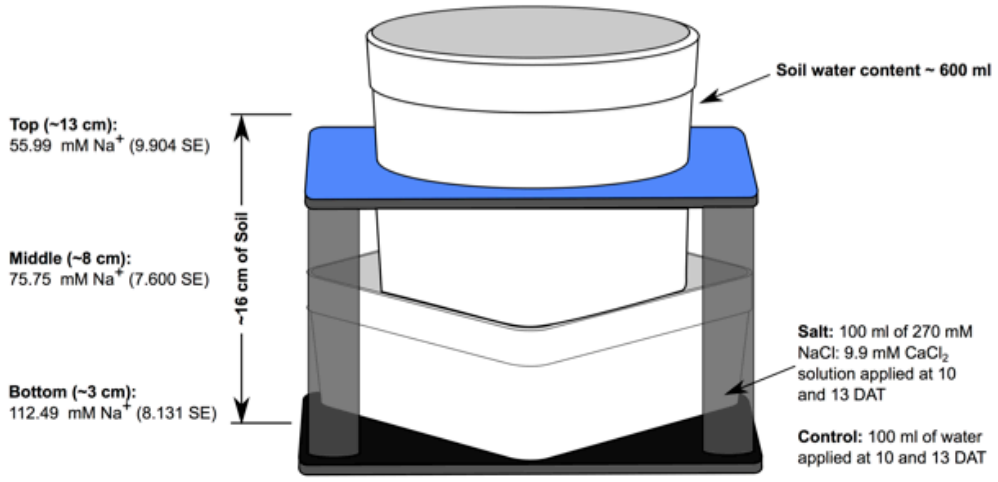


Figure S2. Schematic illustrating NaCl application. The salt treatment was applied by adding 100 ml of NaCl solution (270 mM NaCl: 9.9 mM $CaCl_ß$) the square dish beneath the pot. The treatment was applied in two steps of 45 mM to reach a final concentration of 90 mM, at 10 and 13 DAT. Soil sodium concentration throughout the soil profile is listed on the left. DAT: Days after transplant.

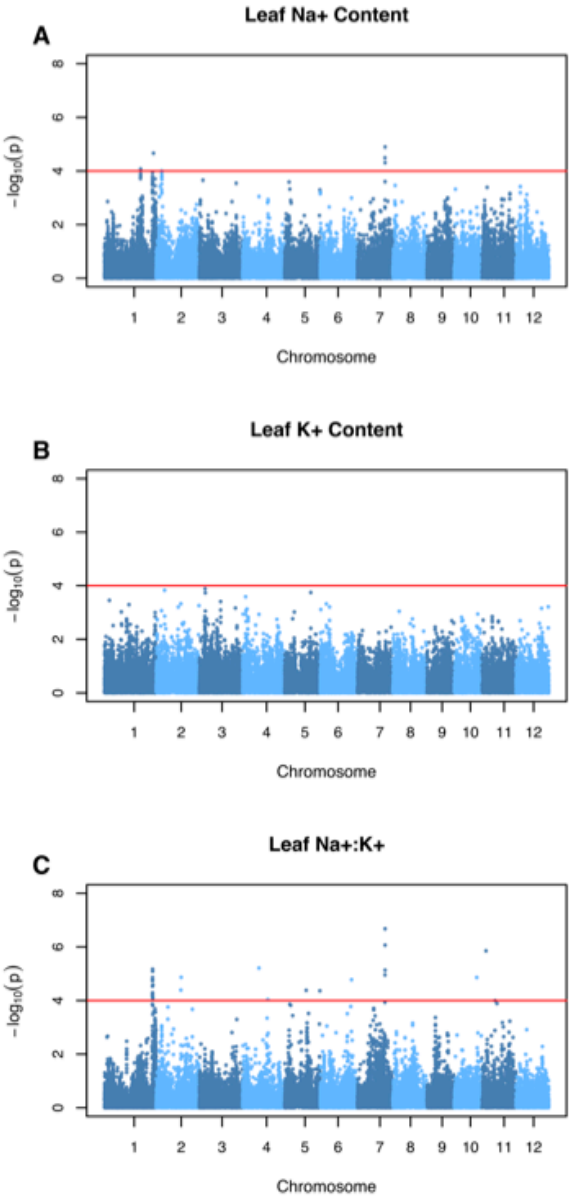


Figure S3. GWA analysis of leaf Na^+ (A), K^+ (B) and Na^+:K^+ content (C). The third leaf was collected at day 15 after 90 mM NaCl. The red horizontal line indicates statistical significance ($p < 10^{-4}$).

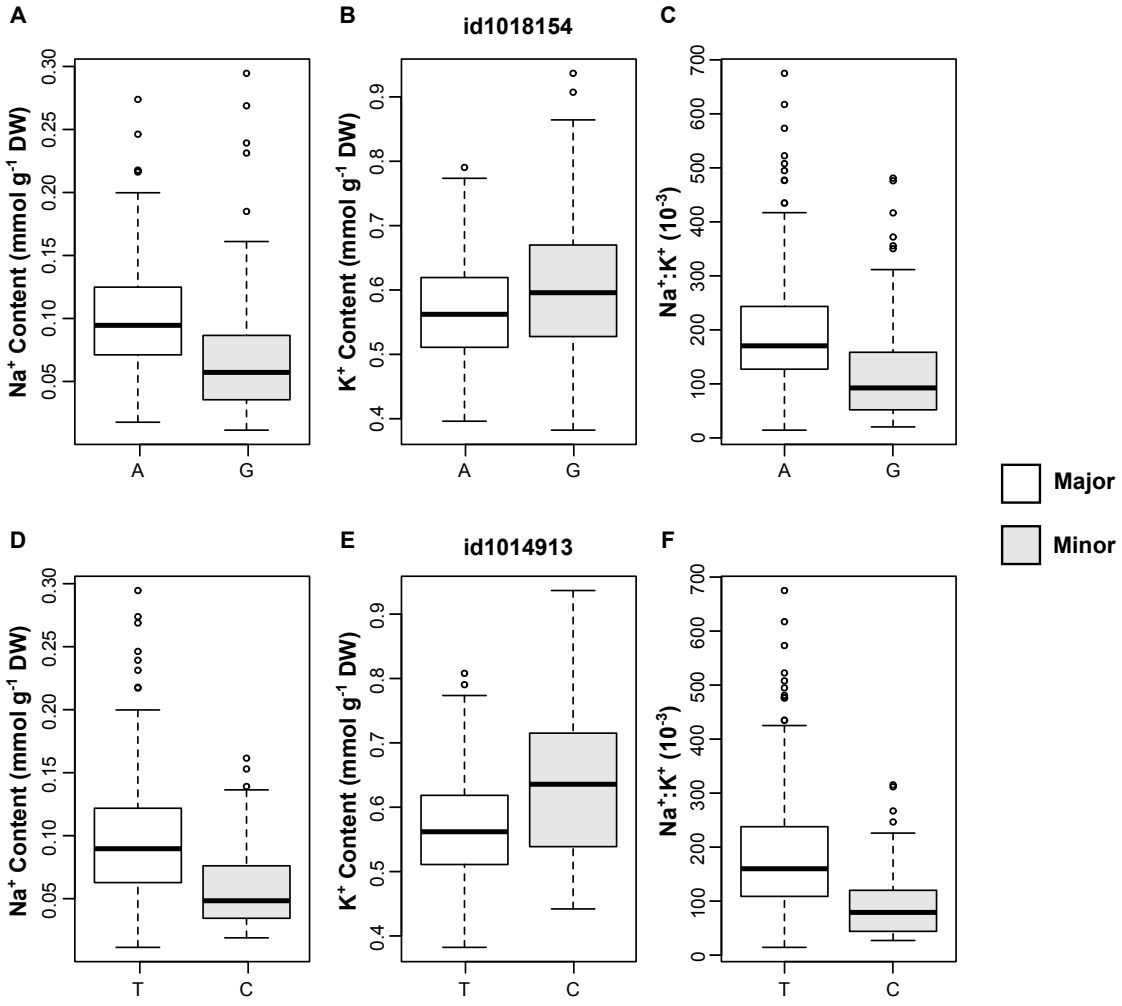


Fig S4: Distribution of leaf ion content observed between allelic groups at id1018154 (A-C) and id1014913 (D-F). (A) The minor allelic group at id1018154 is indicated by "G" where *n*=147, while "A" represents the major allelic group where *n*=173. (B) "C" indicates the minor allelic group at id1014913 where *n*=69, and "T" represents the major allelic group where *n*=263.

Fig S5-18: (A-D) GWA analysis of fluorescence color classes CC.21 (A), CC.31 (B), CC.41 (C) and CC.52 (D) at day 1-14 after 90 mM NaCl. Red horizontal line indicates statistical significance ($p < 10^{-7}$).

Fig S5.

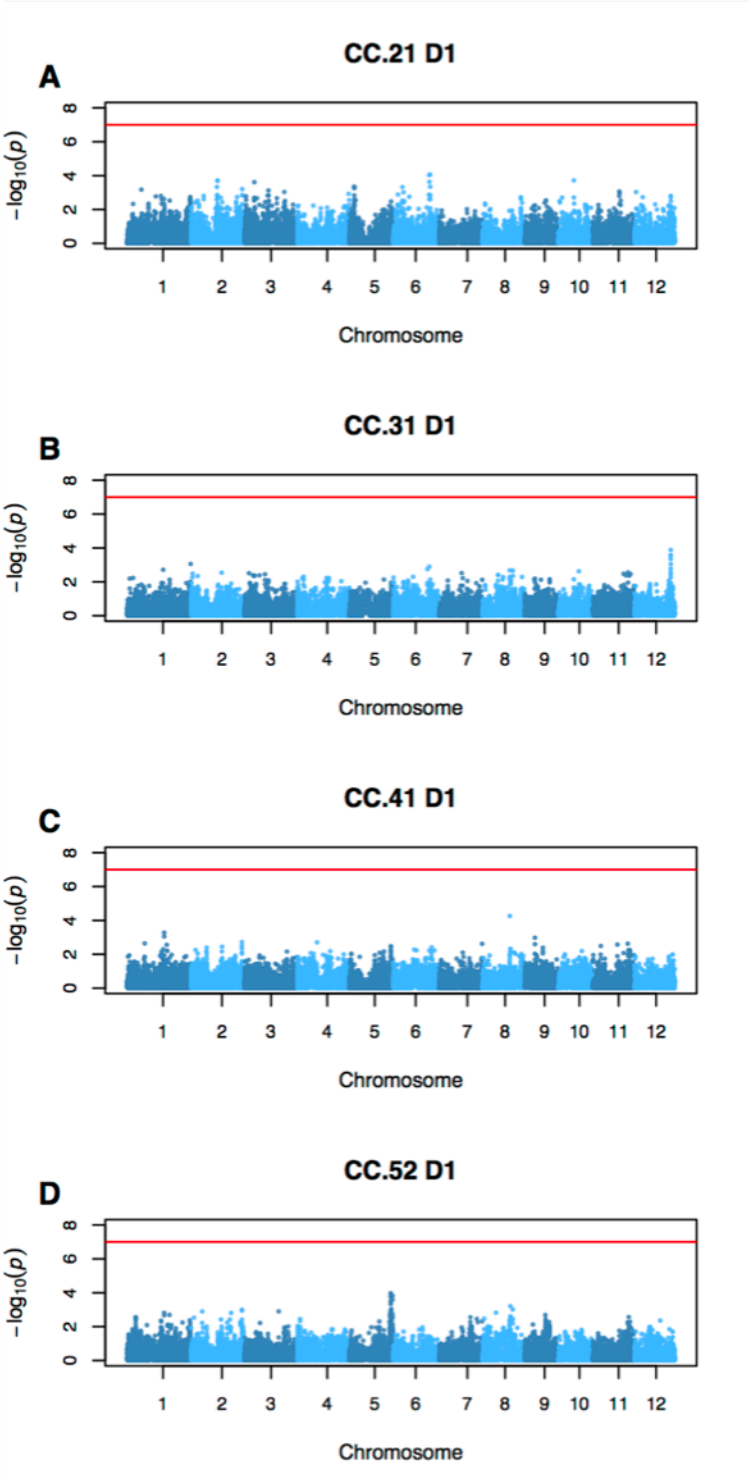

A
CC.21 D1
B
CC.31 D1
C
CC.41 D1
D
CC.52 D1
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S6.

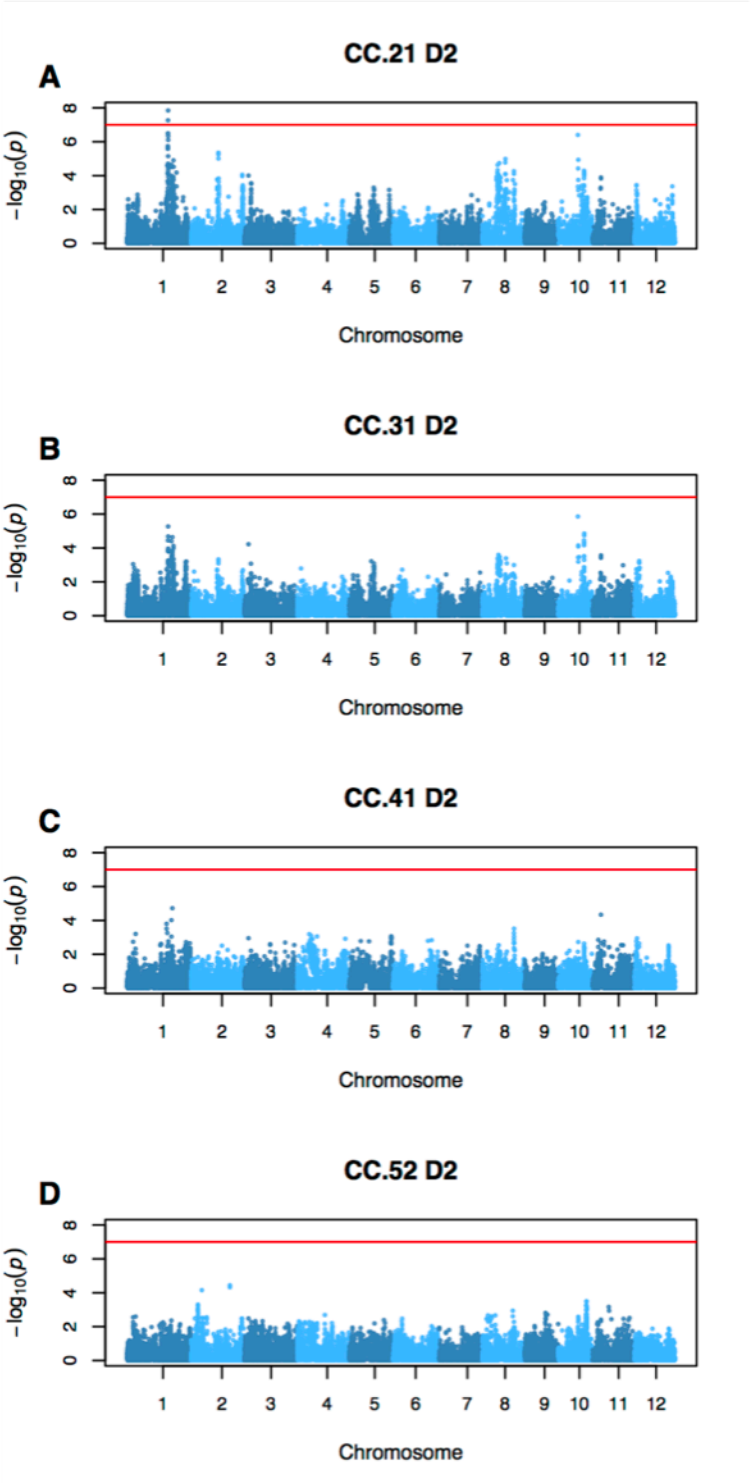

A
CC.21 D2
B
CC.31 D2
C
CC.41 D2
D
CC.52 D2
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S7.

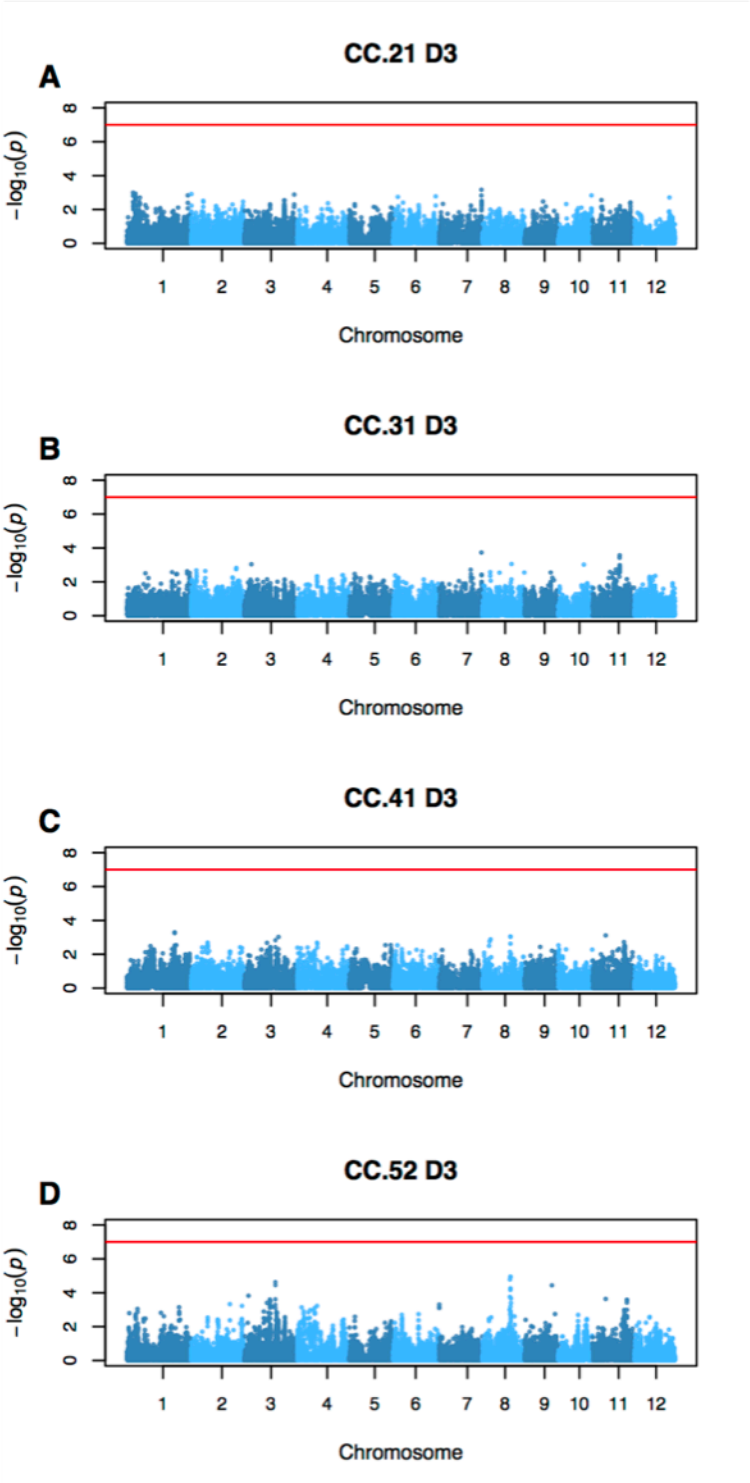

A
CC.21 D3
B
CC.31 D3
C
CC.41 D3
D
CC.52 D3
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S8.

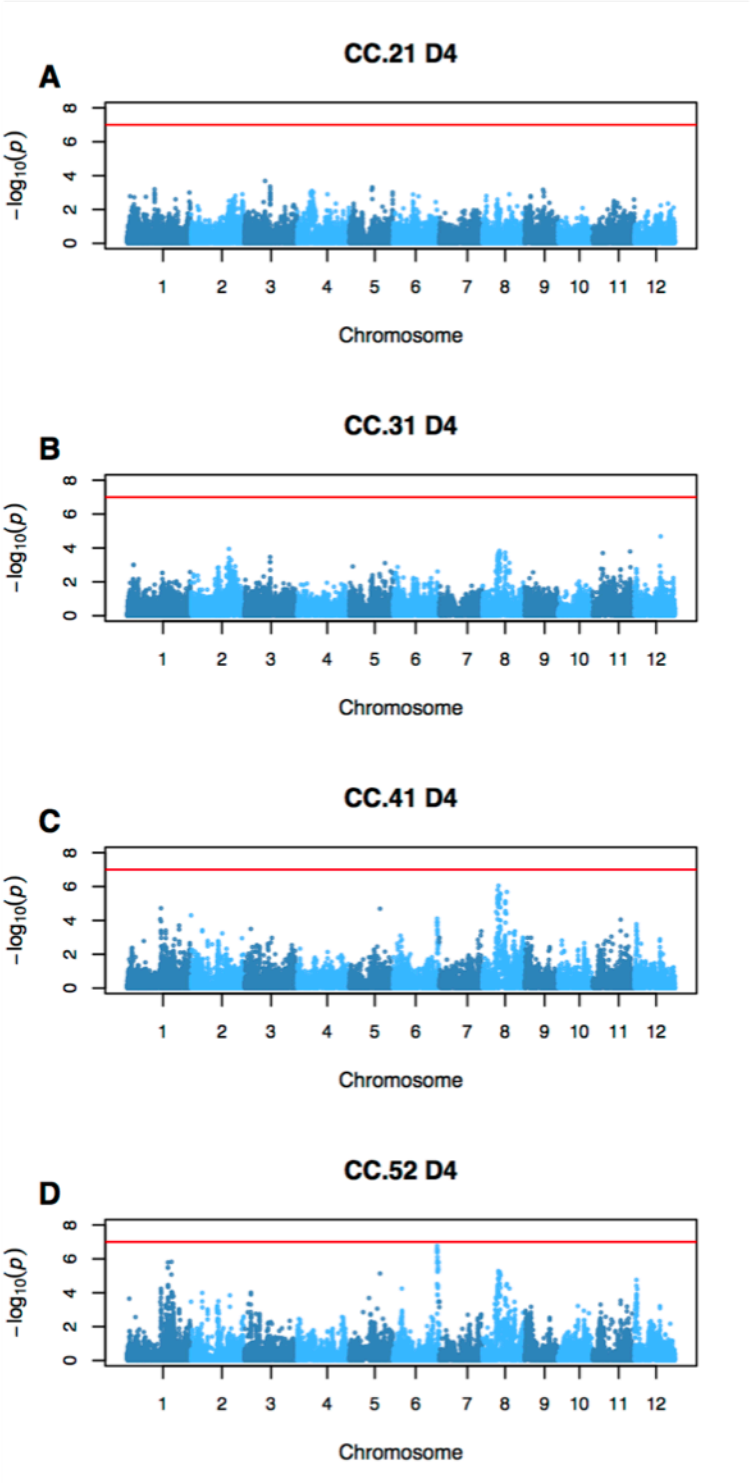

A
CC.21 D4
B
CC.31 D4
C
CC.41 D4
D
CC.52 D4
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S9.

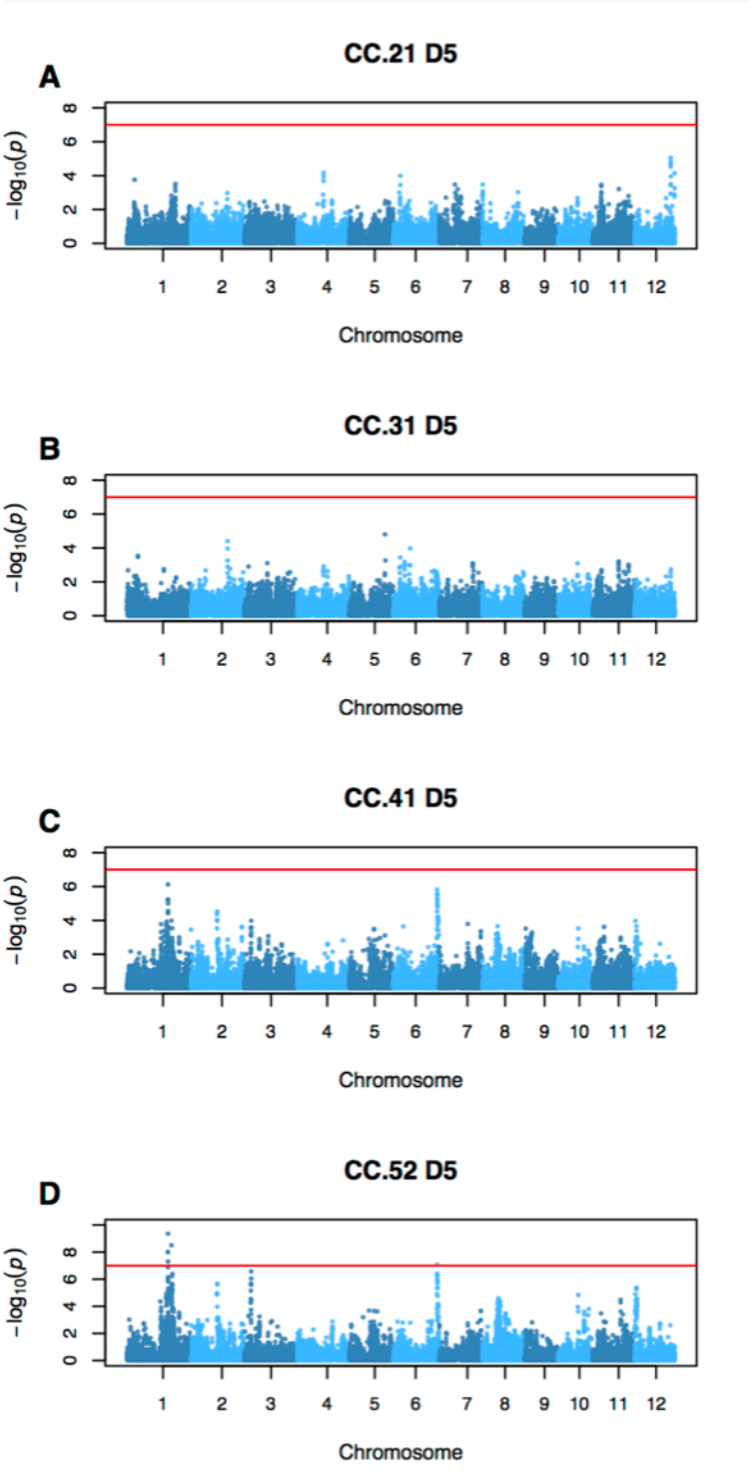

A
CC.21 D5
B
CC.31 D5
C
CC.41 D5
D
CC.52 D5
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S10.

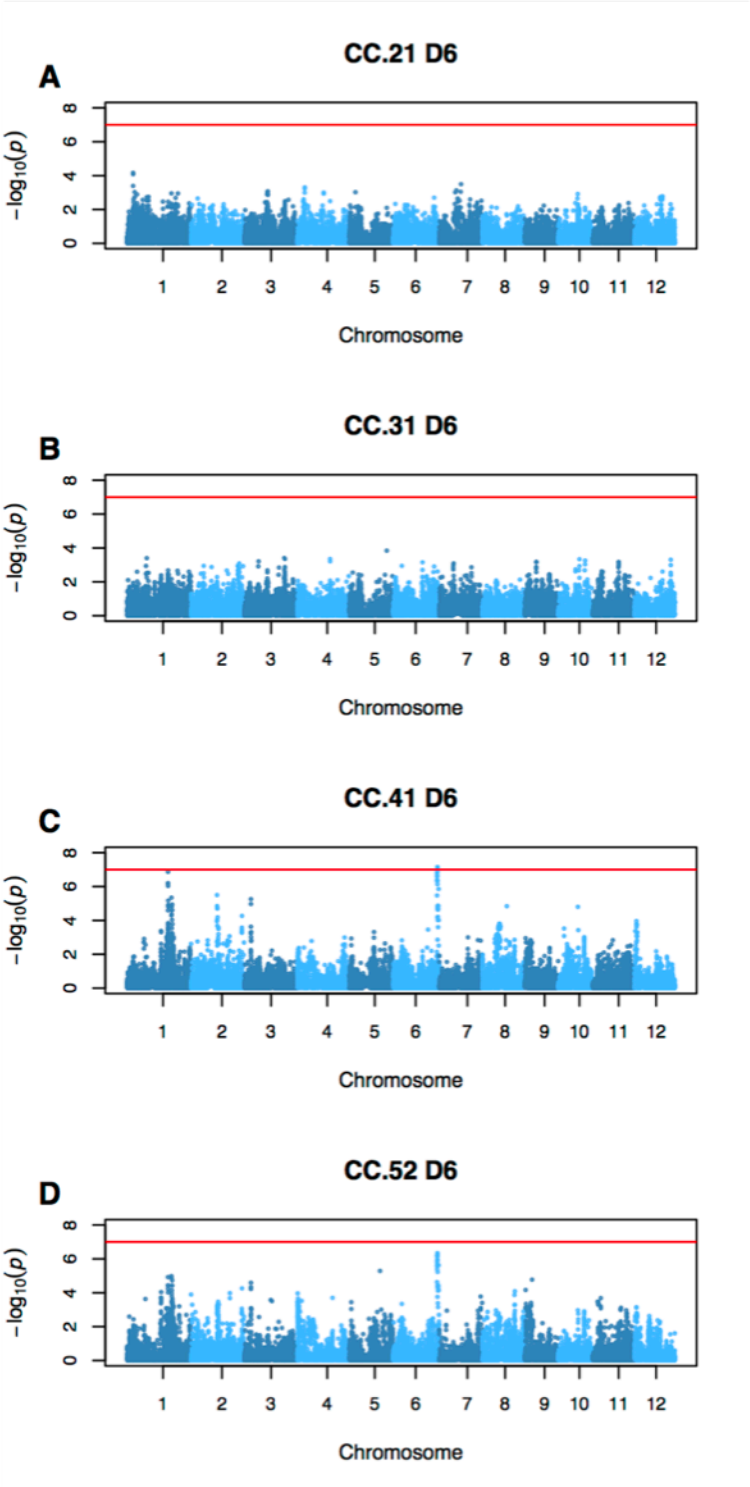

A
CC.21 D6
B
CC.31 D6
C
CC.41 D6
D
CC.52 D6
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S11.

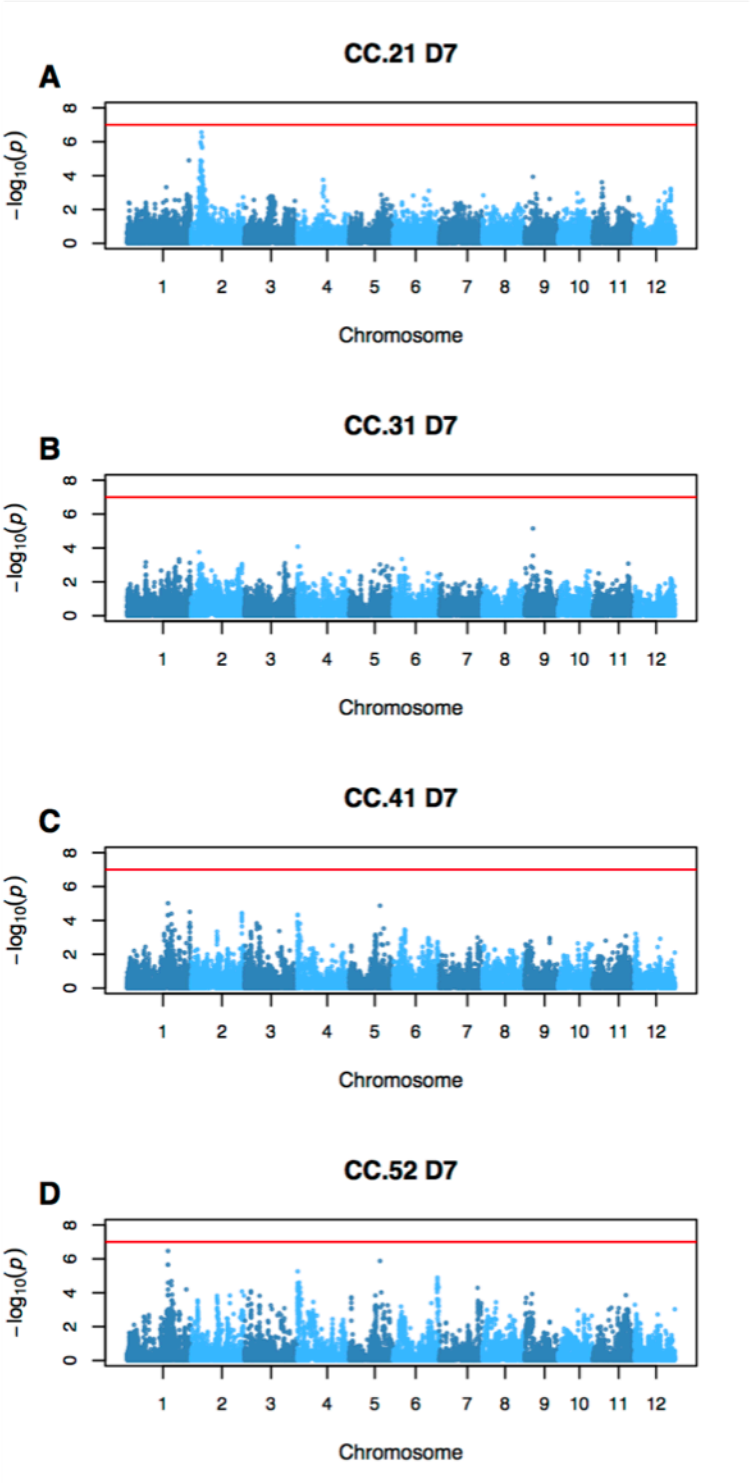

A
CC.21 D7
B
CC.31 D7
C
CC.41 D7
D
CC.52 D7
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S12.

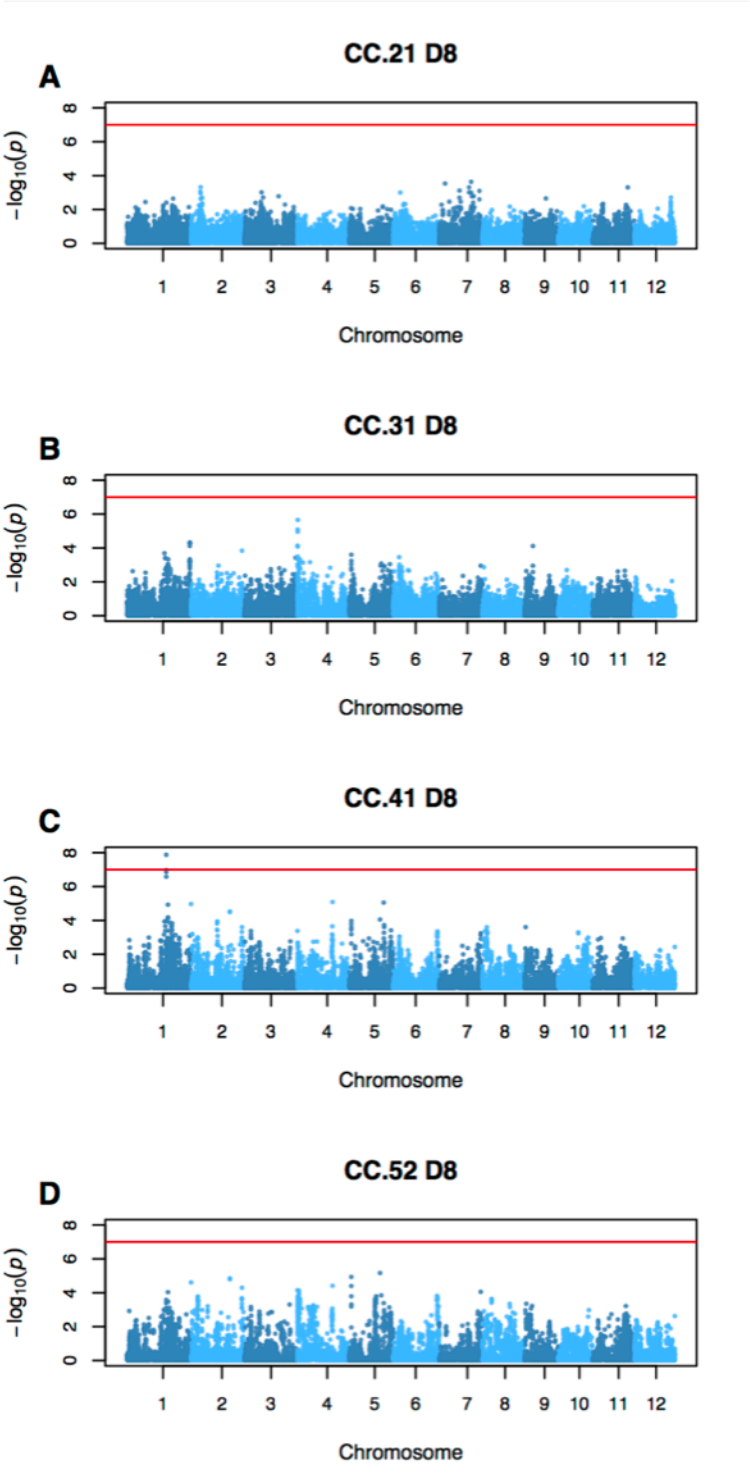

A
CC.21 D8
B
CC.31 D8
C
CC.41 D8
D
CC.52 D8
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S13.

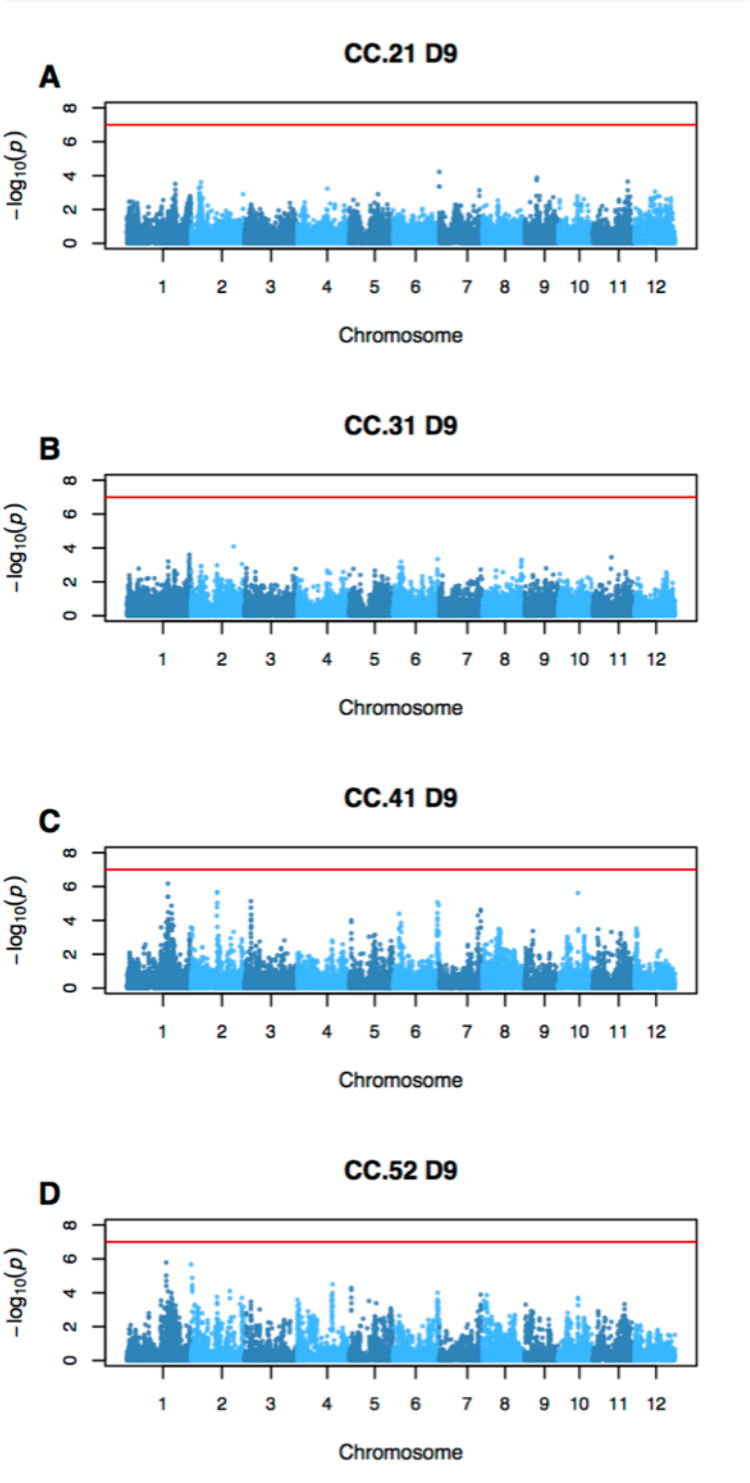

A
CC.21 D9
B
CC.31 D9
C
CC.41 D9
D
CC.52 D9
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S14.

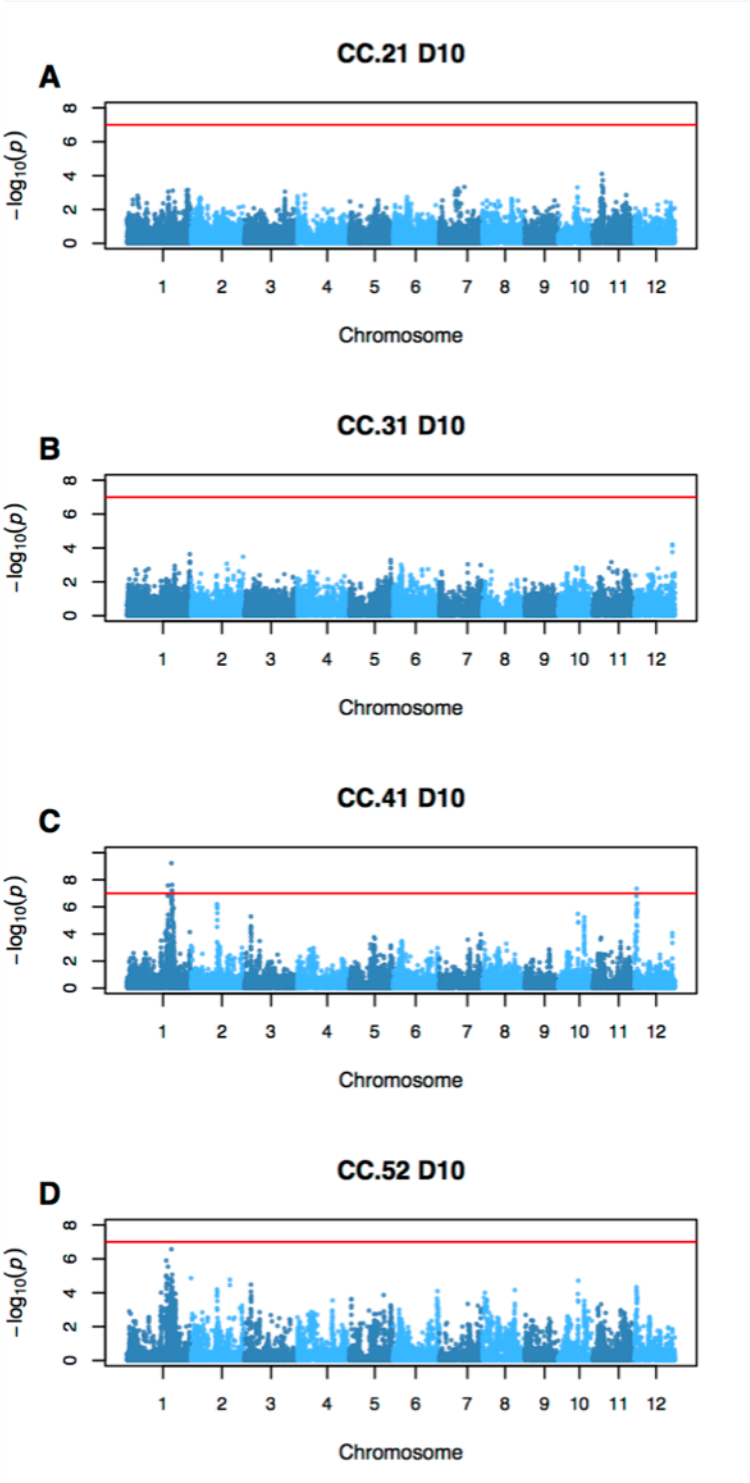

A
CC.21 D10
B
CC.31 D10
C
CC.41 D10
D
CC.52 D10
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S15.

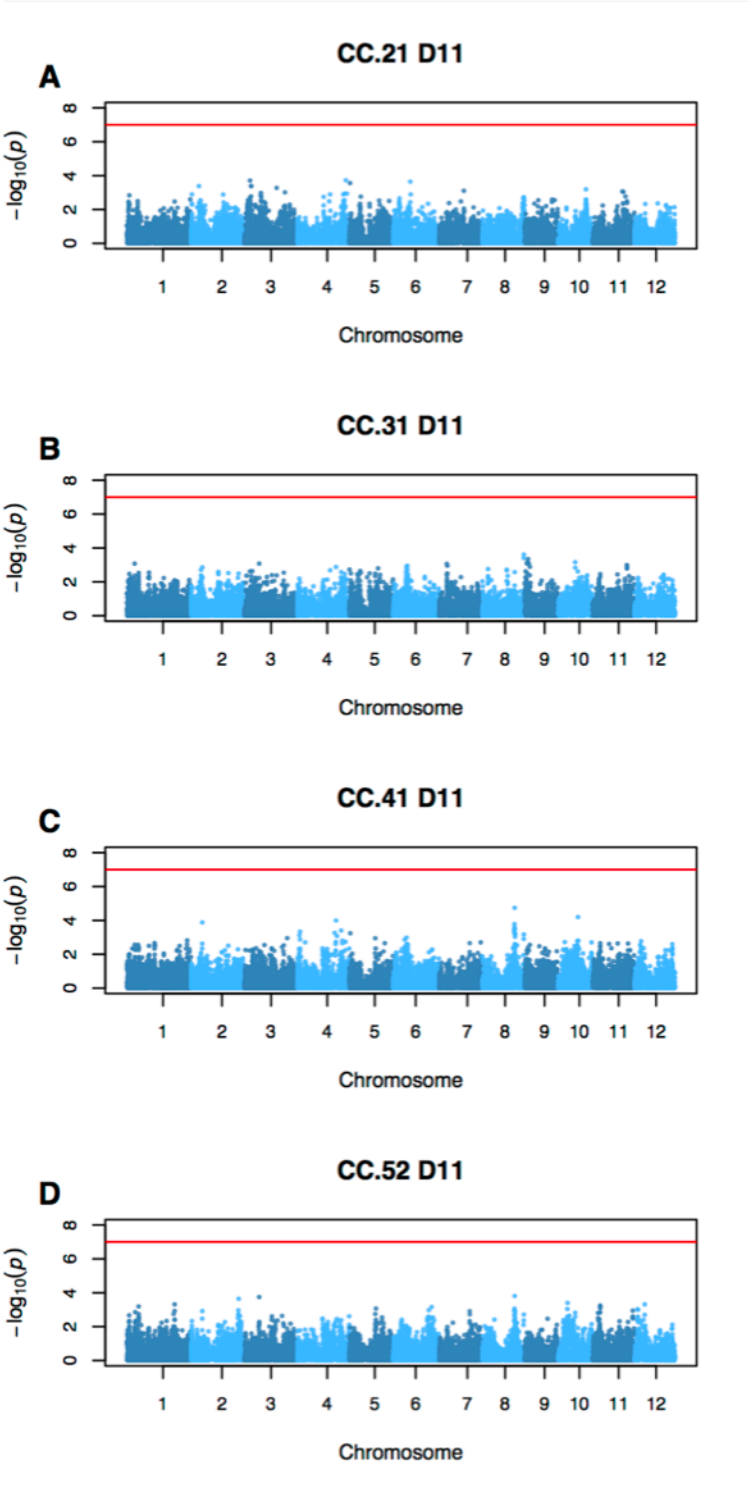

A
CC.21 D11
B
CC.31 D11
C
CC.41 D11
D
CC.52 D11
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S16.

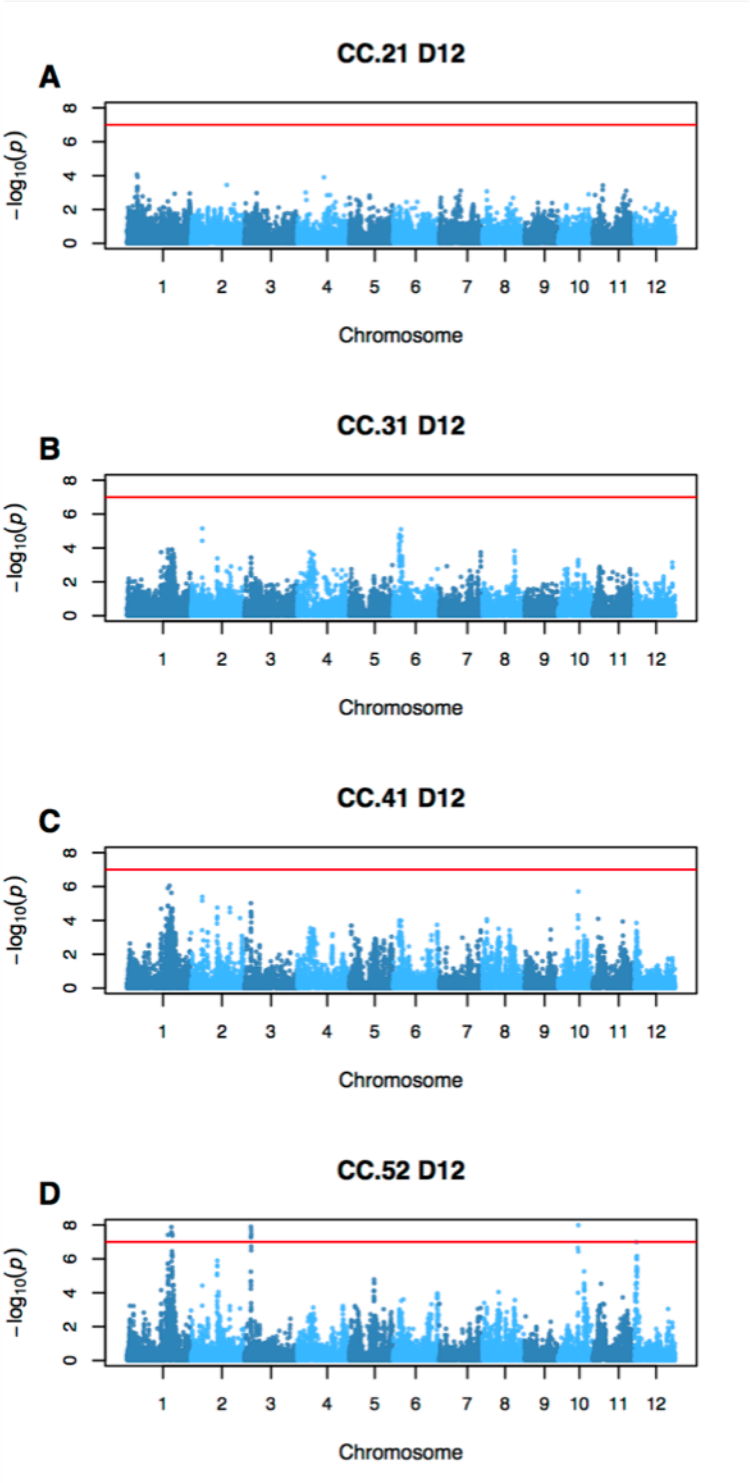

A
CC.21 D12
B
CC.31 D12
C
CC.41 D12
D
CC.52 D12
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S17.

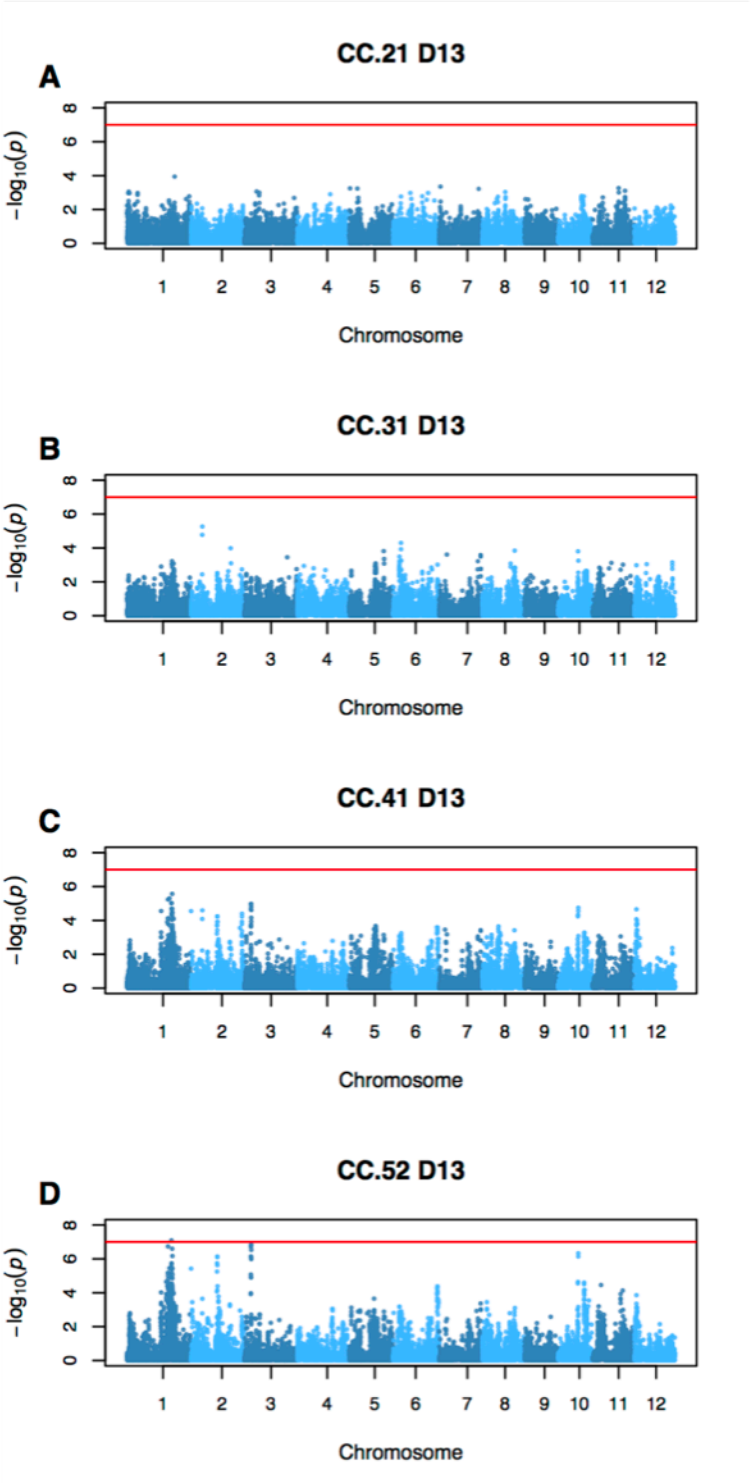

A
CC.21 D13
B
CC.31 D13
C
CC.41 D13
D
CC.52 D13
−log10(p)
0 2 4 6 8
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Fig S18.

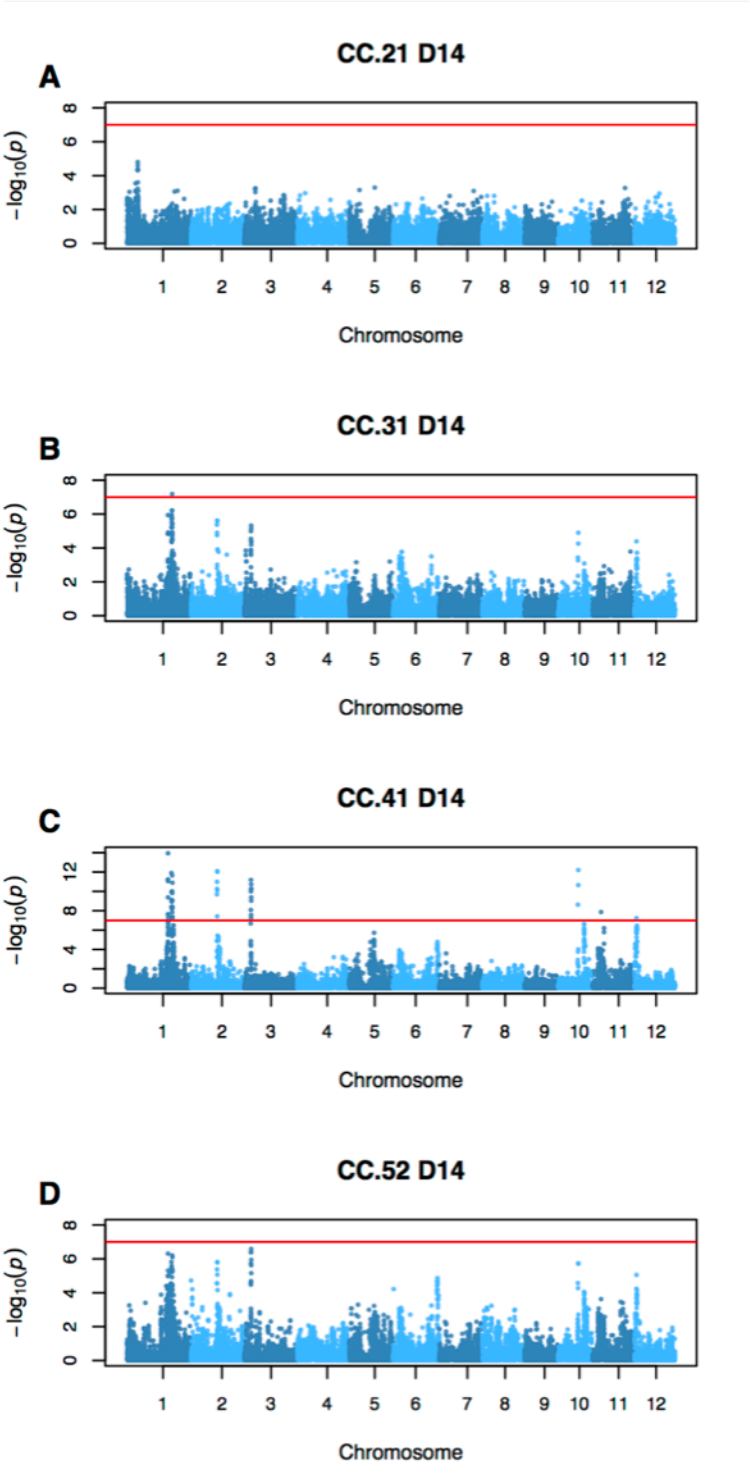
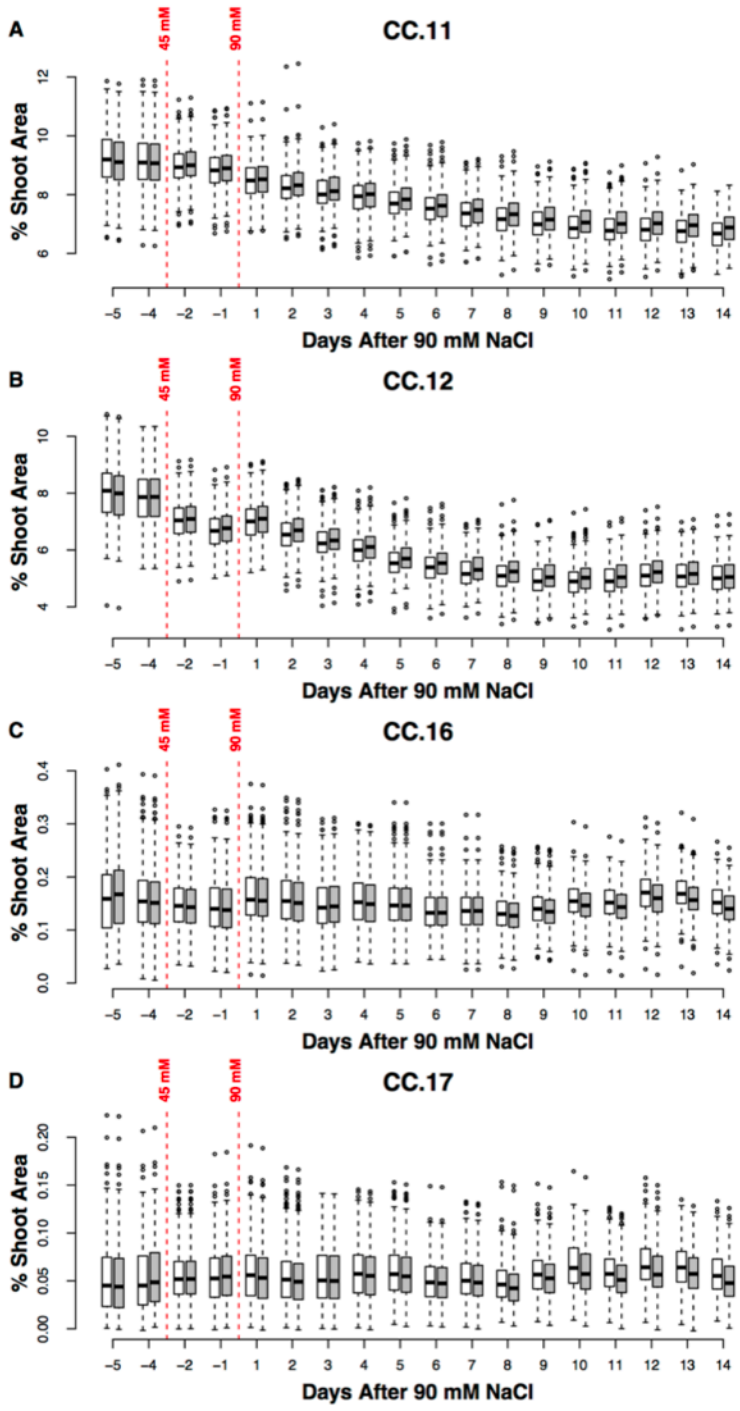
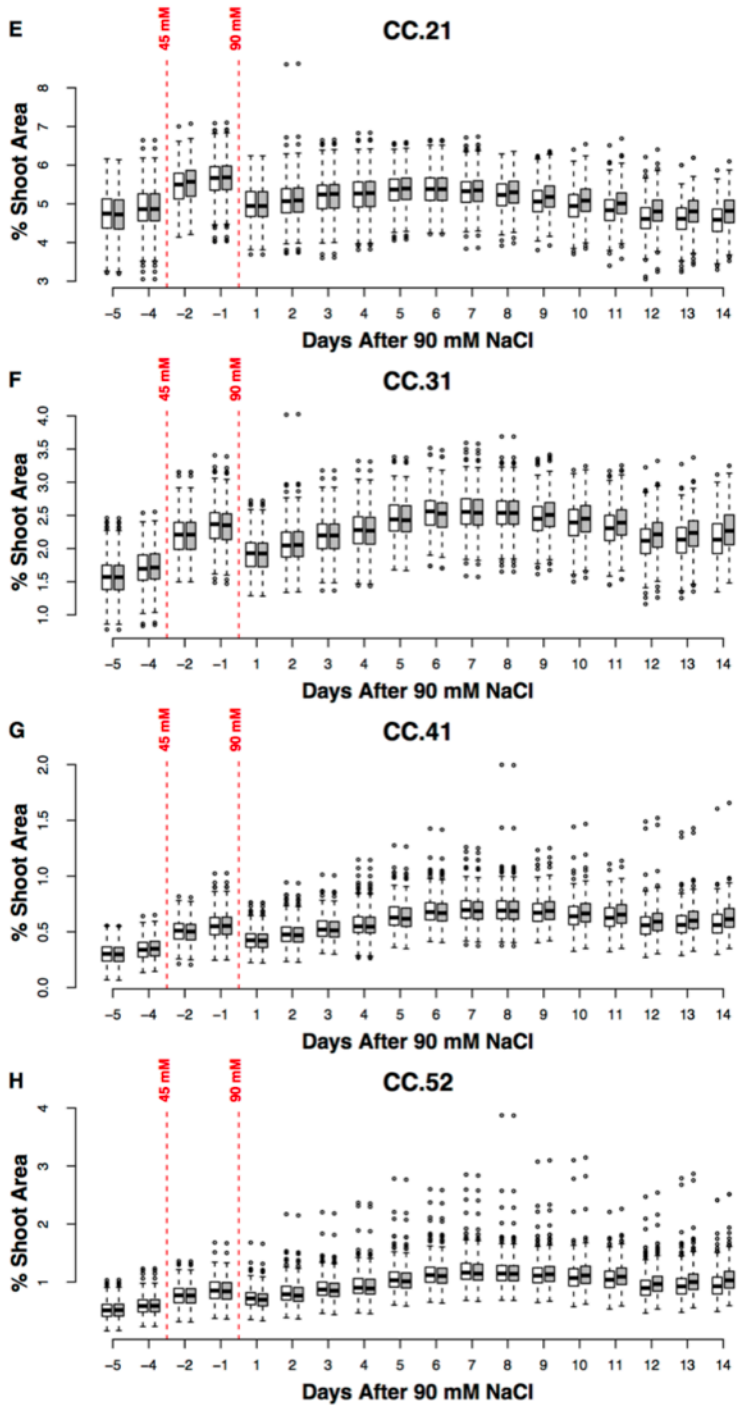
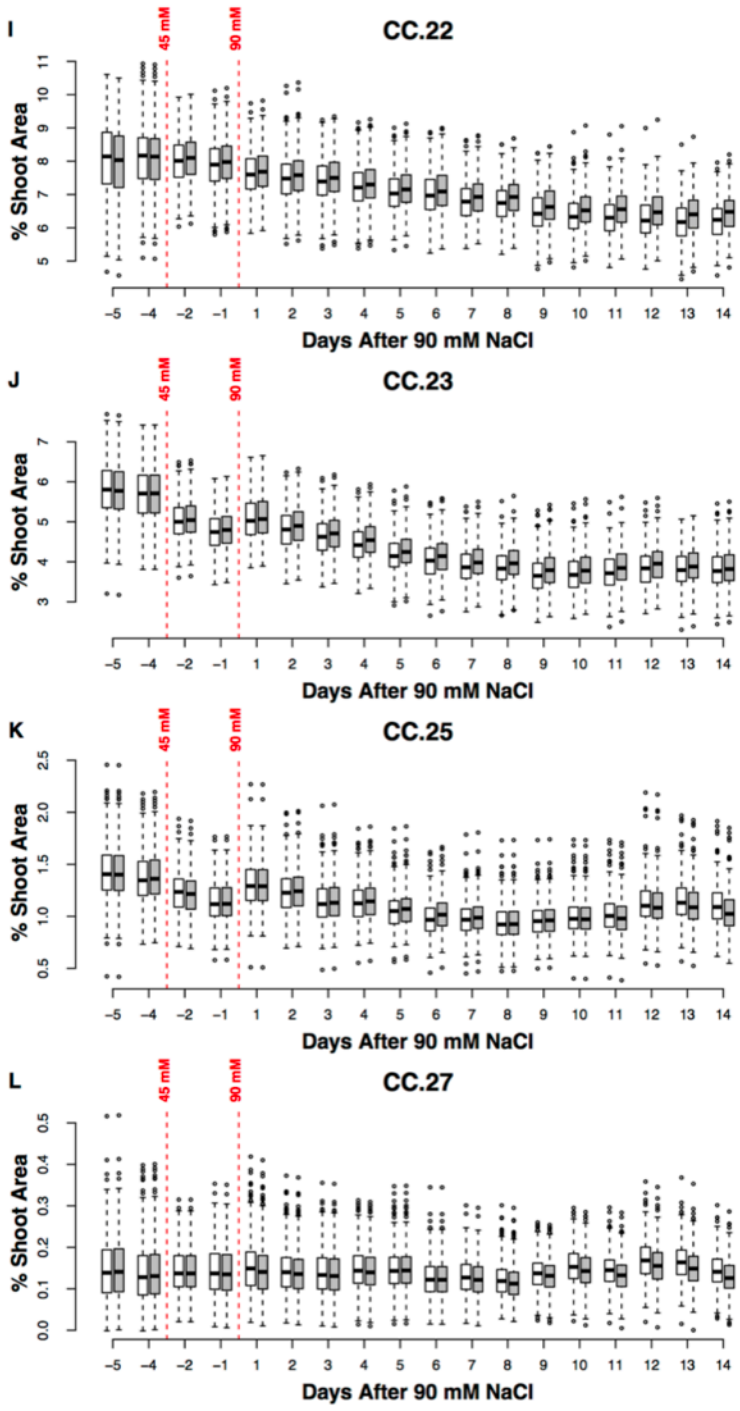
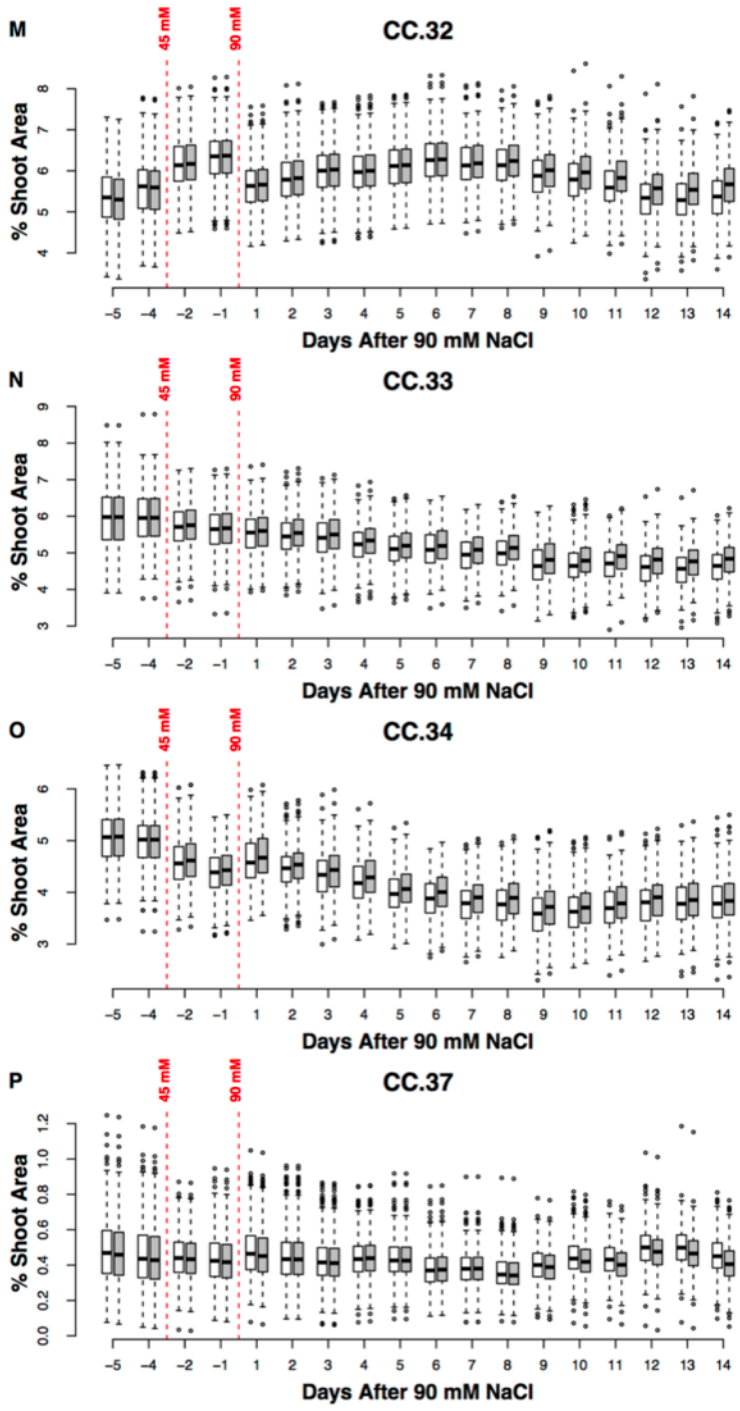

A
CC.21 D14
$-\log_{10}(p)$
Chromosome
B
CC.31 D14
$-\log_{10}(p)$
Chromosome
C
CC.41 D14
$-\log_{10}(p)$
Chromosome
D
CC.52 D14
$-\log_{10}(p)$
Chromosome

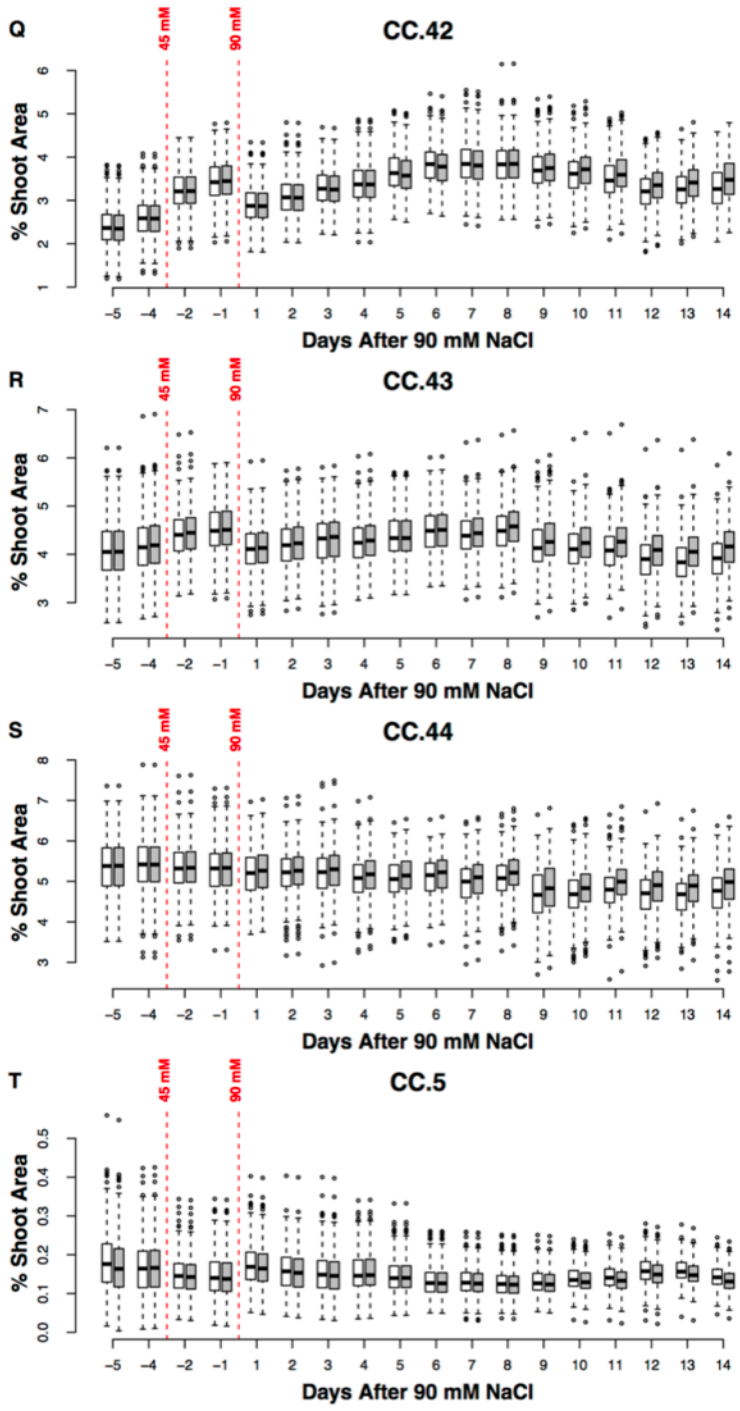
Fig S19: (A-AF) Trajectories of the 32 salinity-responsive fluorescence color classes. Control plants are indicated by white bars, while plants in saline conditions are indicated by grey bars. Values are expressed as a percentage of total pixels from two side view fluorescence images.

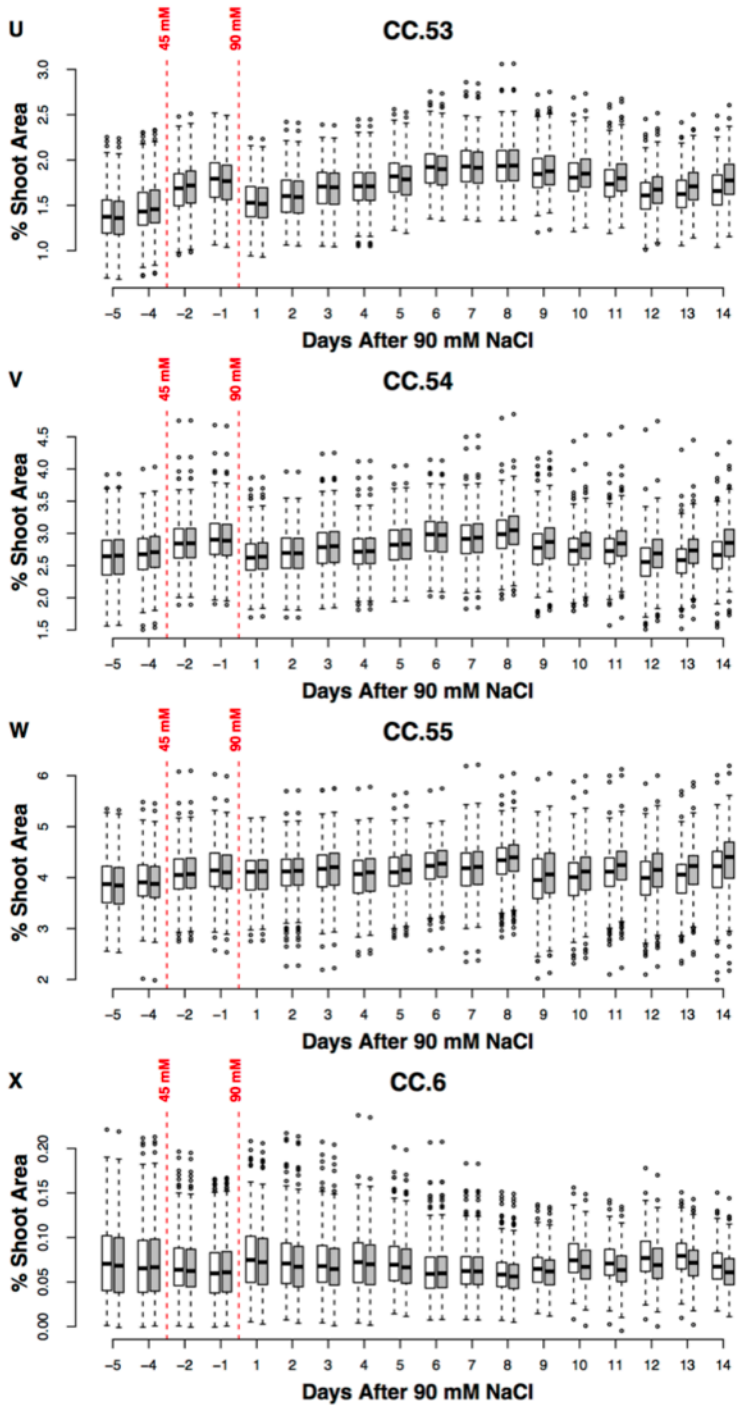

A
CC.11
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
B
CC.12
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
C
CC.16
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
D
CC.17
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl

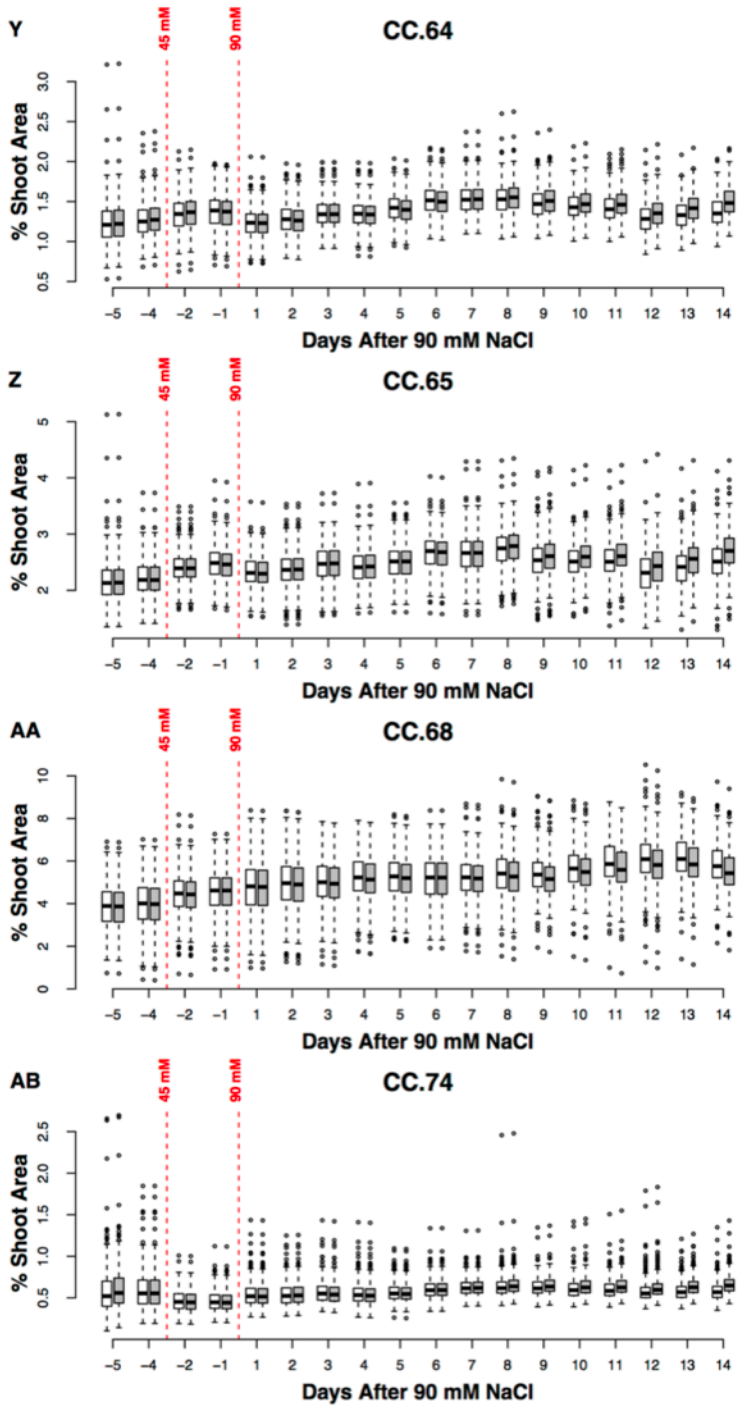

E
CC.21
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
F
CC.31
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
G
CC.41
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
H
CC.52
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl


I
CC.22
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
J
CC.23
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
K
CC.25
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
L
CC.27
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl


M
CC.32
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
N
CC.33
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
O
CC.34
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
P
CC.37
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl


Q
CC.42
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
R
CC.43
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
S
CC.44
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
T
CC.5
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl


U
CC.53
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
V
CC.54
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
W
CC.55
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
X
CC.6
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl


Y
CC.64
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
Z
CC.65
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
AA
CC.68
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
AB
CC.74
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl

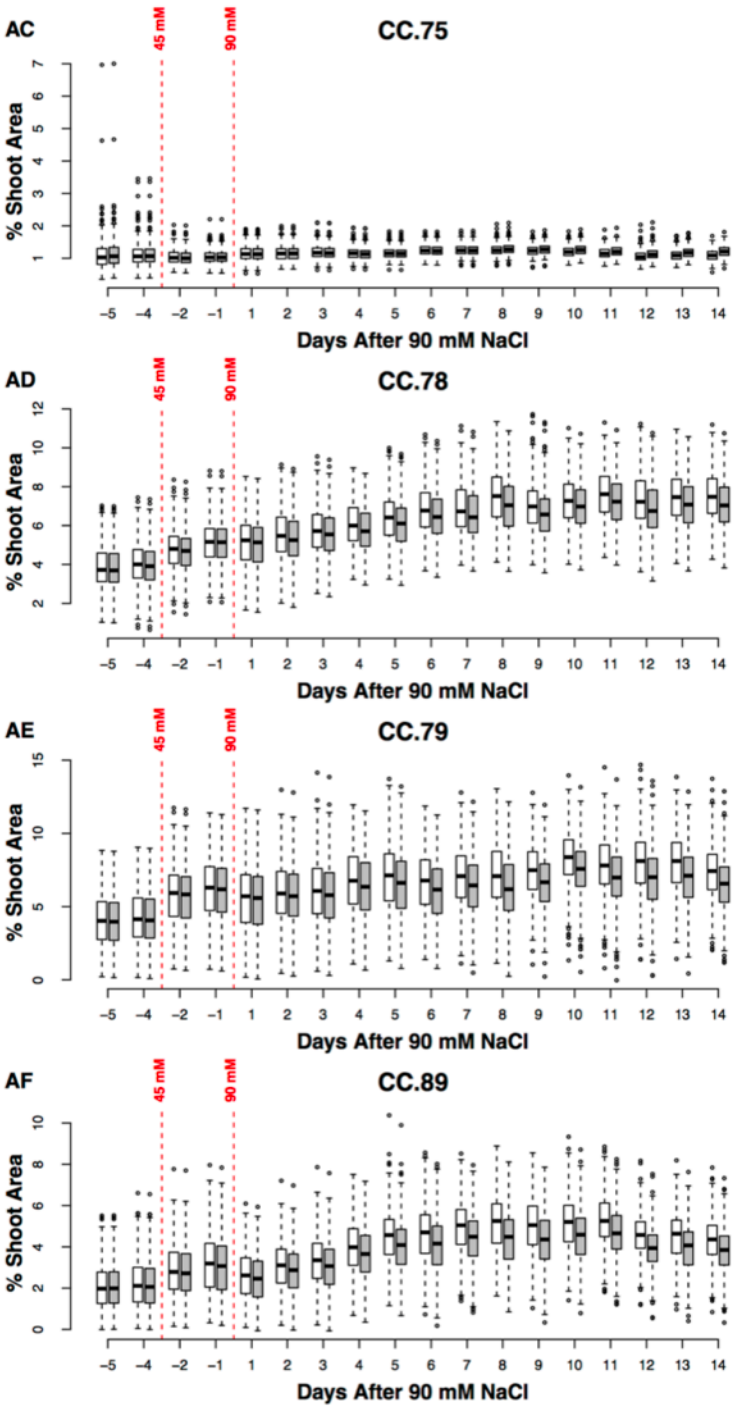
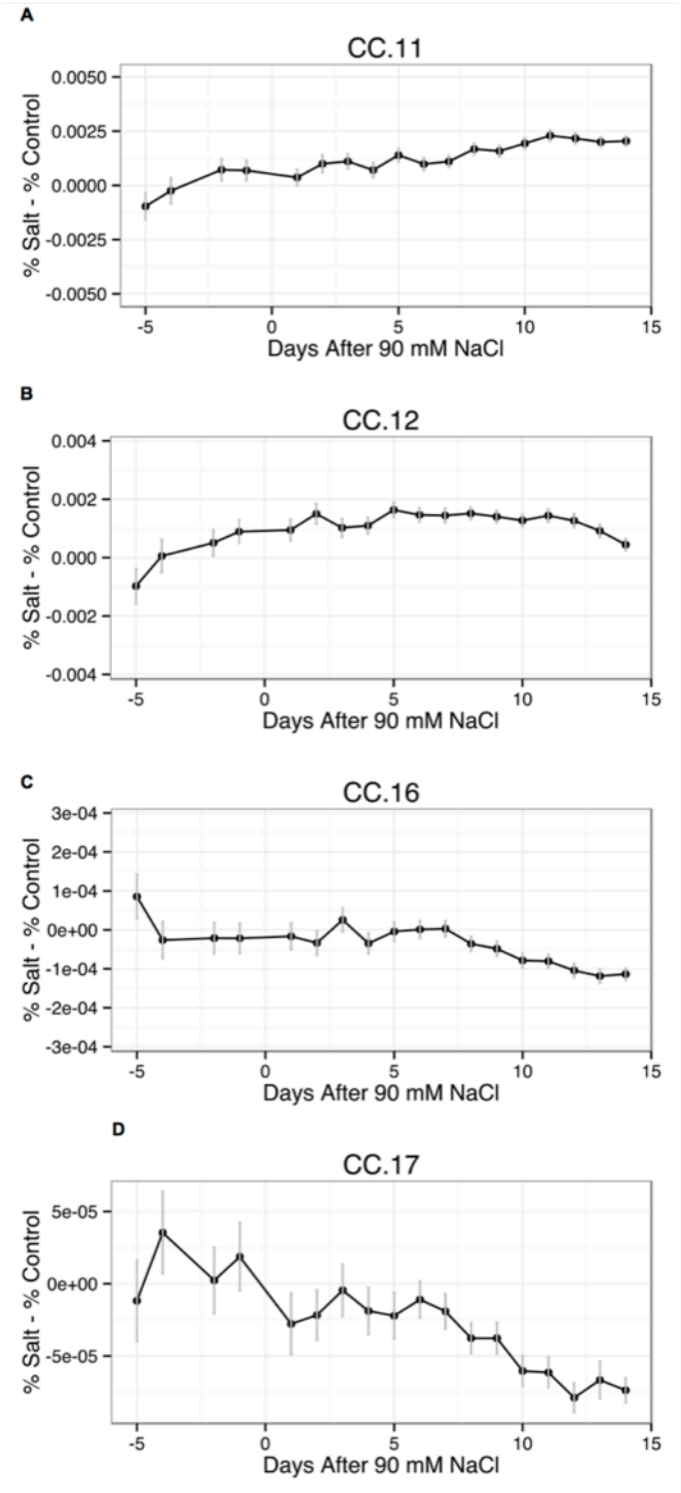
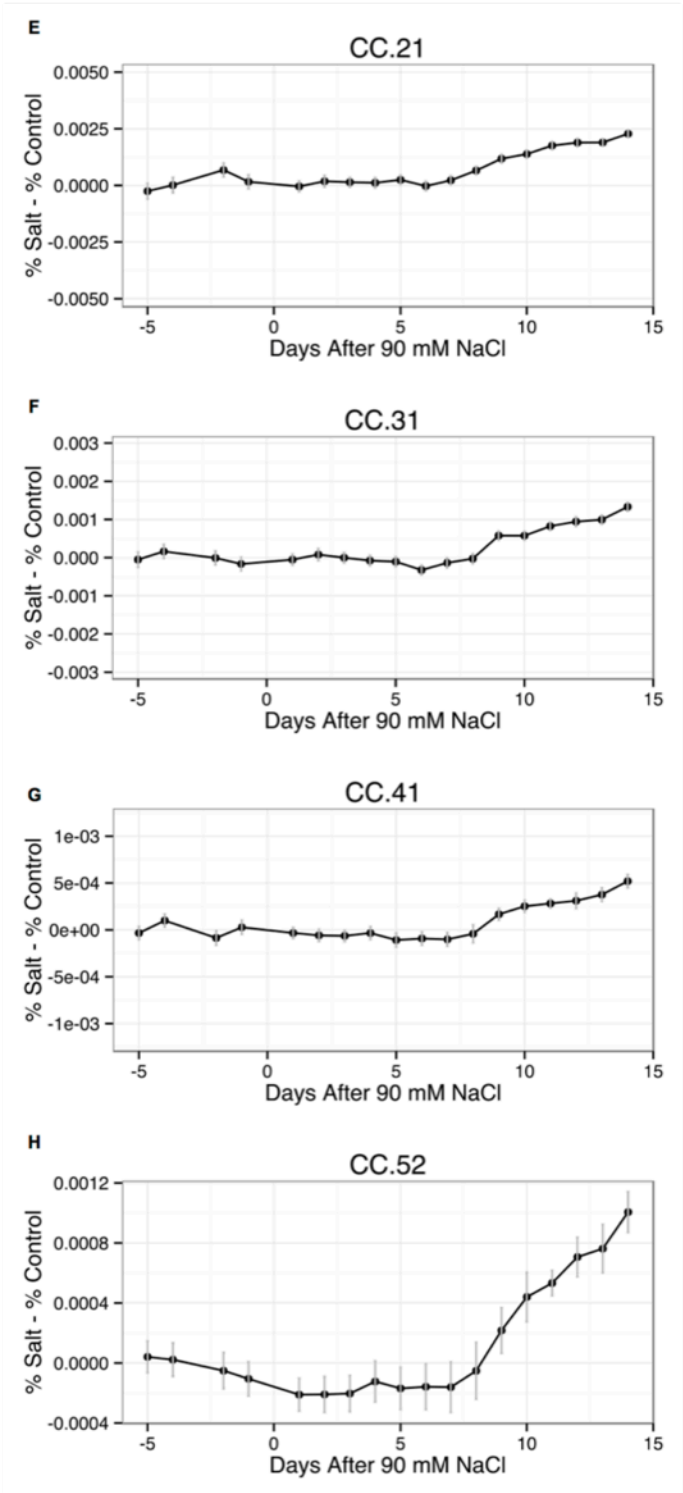
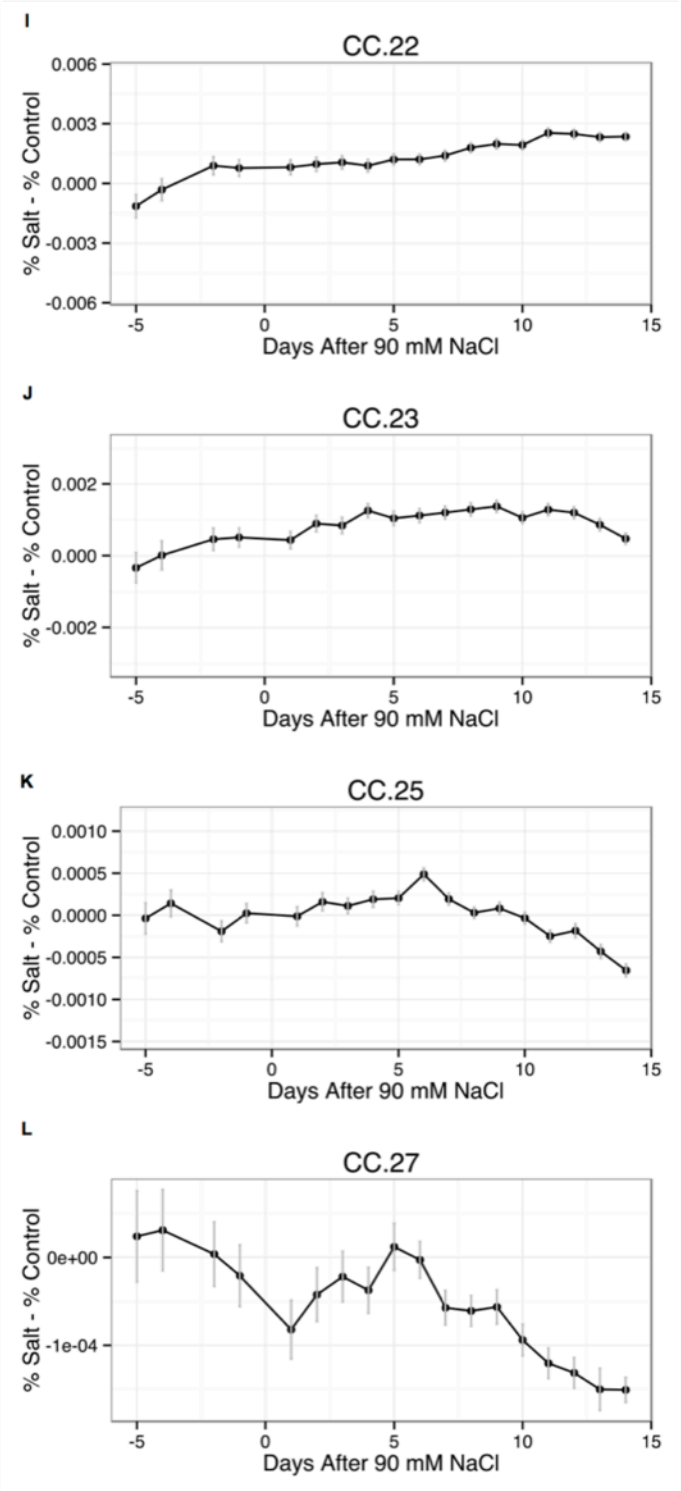
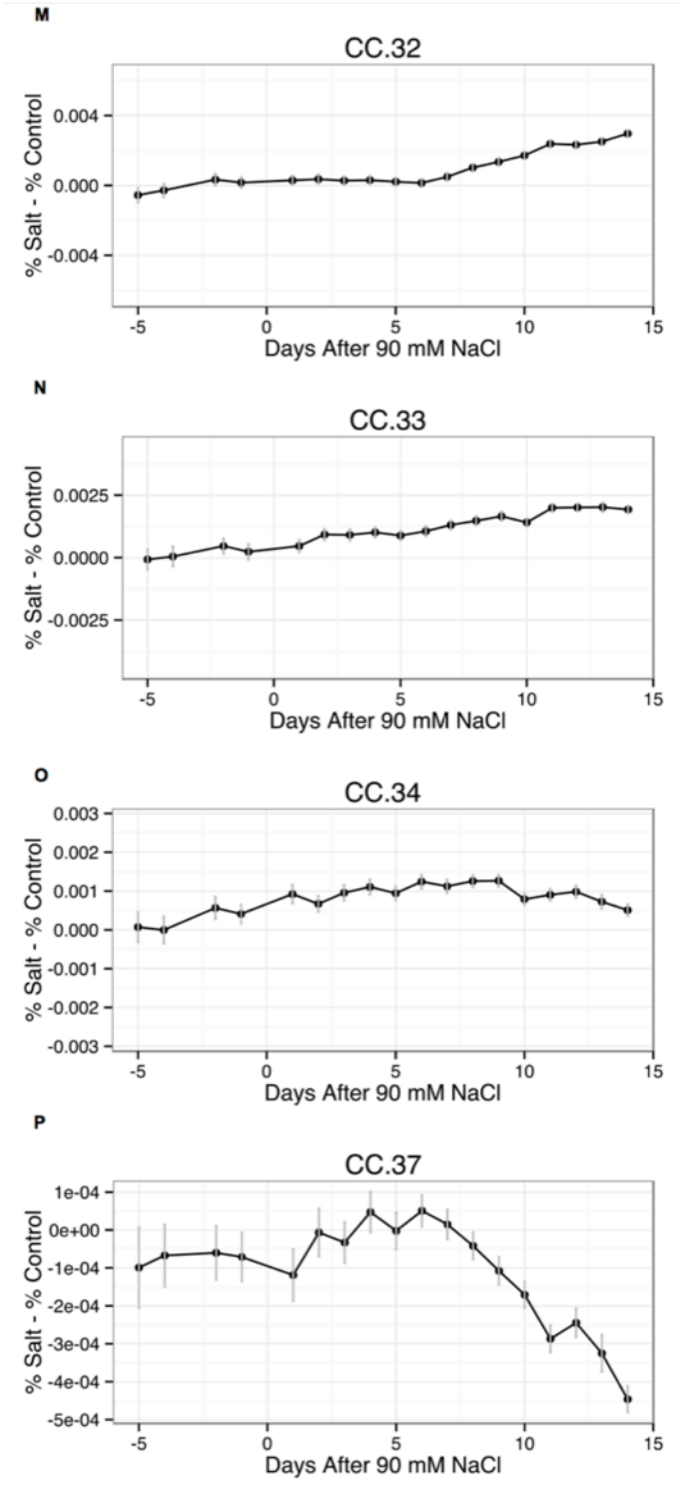

AC
CC.75
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
AD
CC.78
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
AE
CC.79
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl
AF
CC.89
45 mM
90 mM
% Shoot Area
Days After 90 mM NaCl

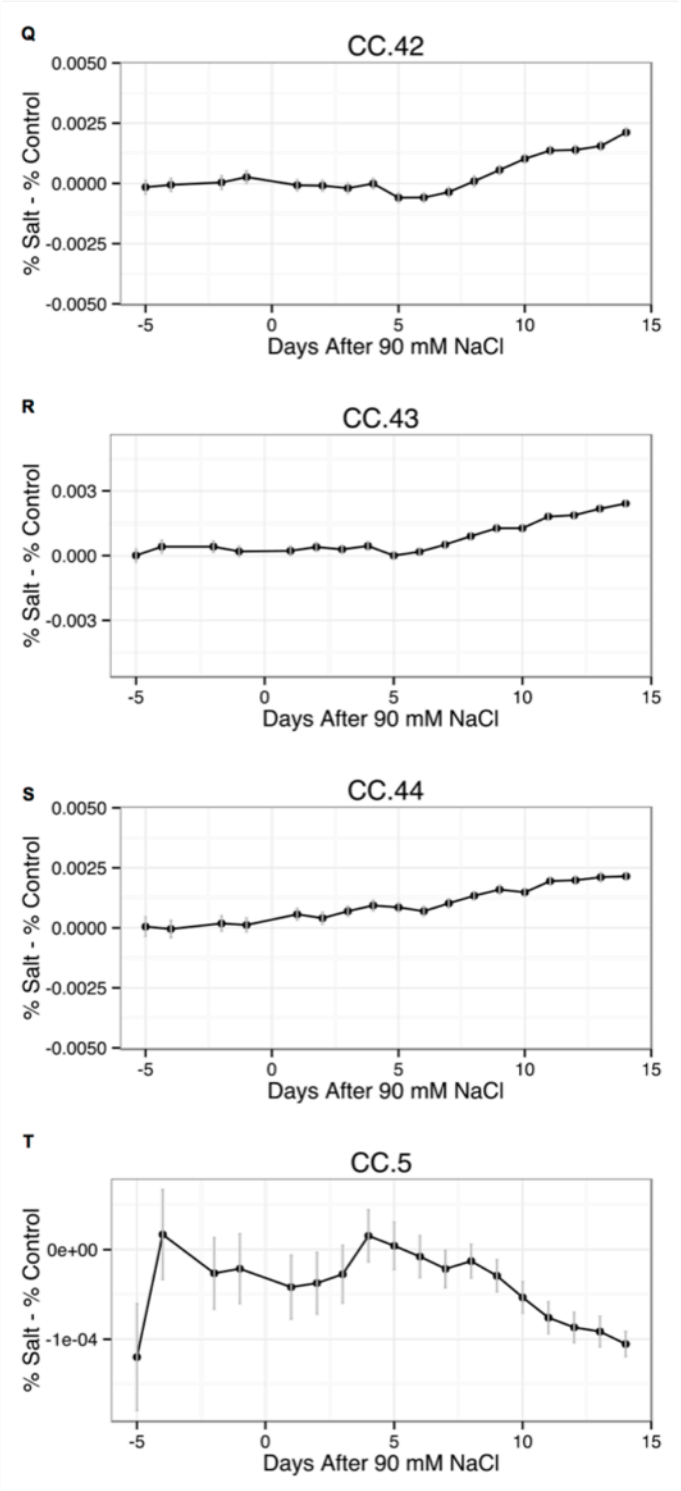
Fig S20: (A-AF) Mean temporal trends of the 32 salinity-responsive fluorescence color classes. For each color class and accession we calculated the salinity response, which is defined as the percent of total pixels from two side-view images accounted for by the color class in saline conditions minus the percent of total pixels accounted by the class in control conditions. Error bars represent standard error.

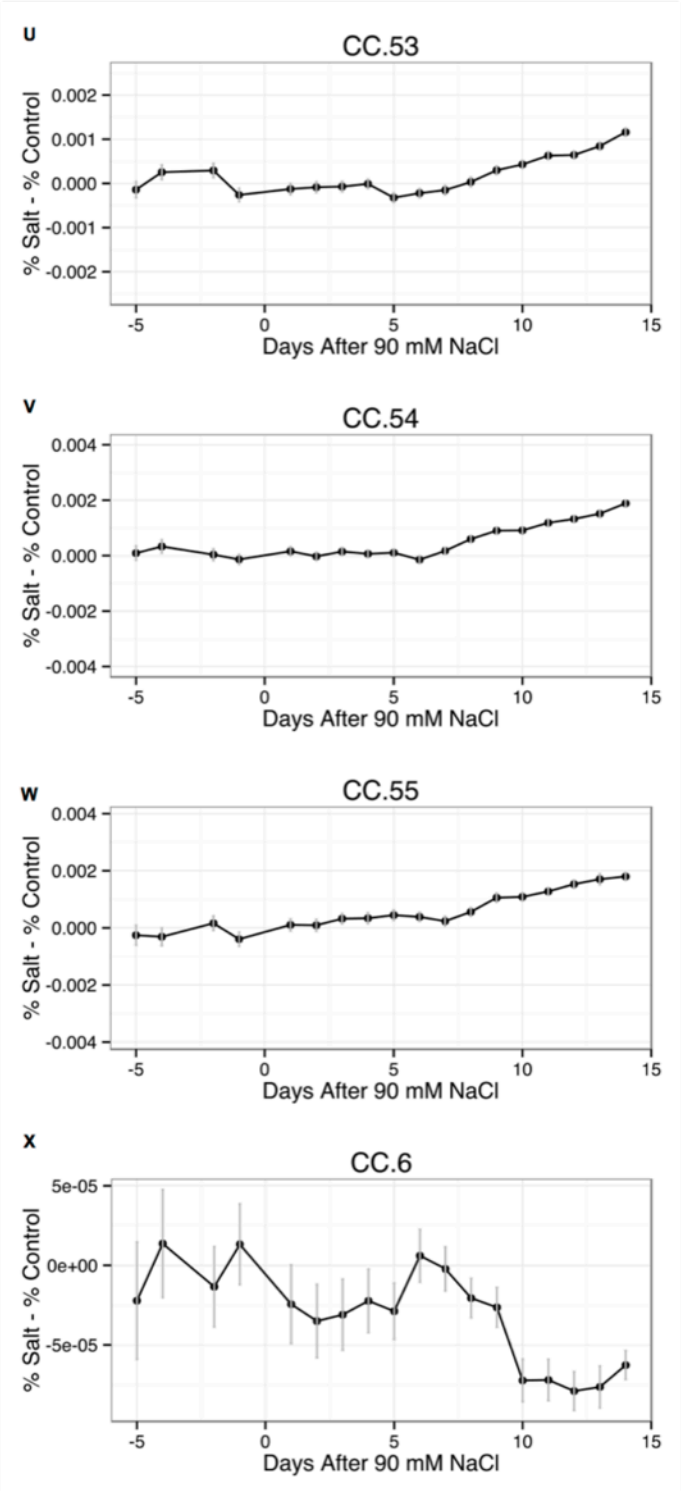

A
CC.11
% Salt - % Control
0.0050
0.0025
0.0000
-0.0025
-0.0050
-5
0
5
10
15
Days After 90 mM NaCl
B
CC.12
% Salt - % Control
0.004
0.002
0.000
-0.002
-0.004
-5
0
5
10
15
Days After 90 mM NaCl
C
CC.16
% Salt - % Control
3e-04
2e-04
1e-04
0e+00
-1e-04
-2e-04
-3e-04
-5
0
5
10
15
Days After 90 mM NaCl
D
CC.17
% Salt - % Control
5e-05
0e+00
-5e-05
-5
0
5
10
15
Days After 90 mM NaCl

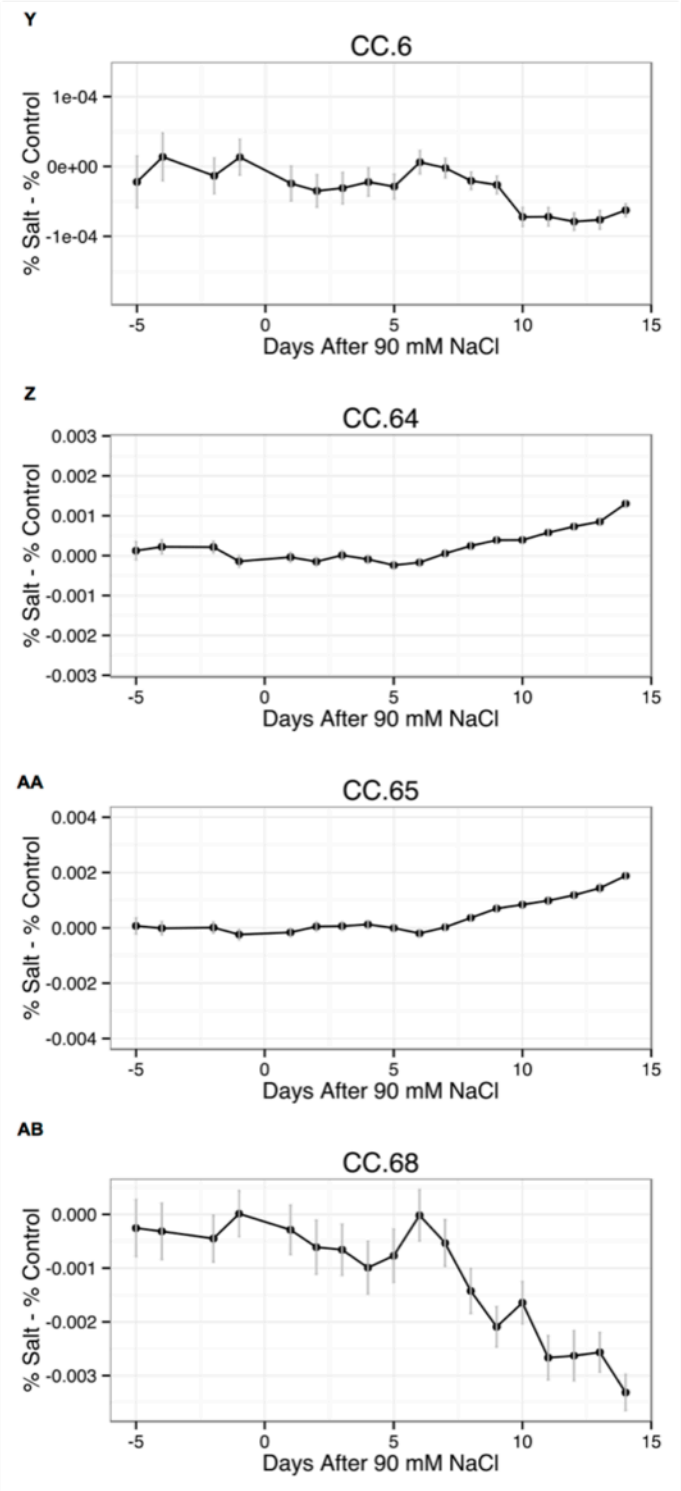

E
CC.21
0.0050
0.0025
0.0000
-0.0025
-0.0050
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
F
CC.31
0.003
0.002
0.001
0.000
-0.001
-0.002
-0.003
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
G
CC.41
1e-03
5e-04
0e+00
-5e-04
-1e-03
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
H
CC.52
0.0012
0.0008
0.0004
0.0000
-0.0004
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl

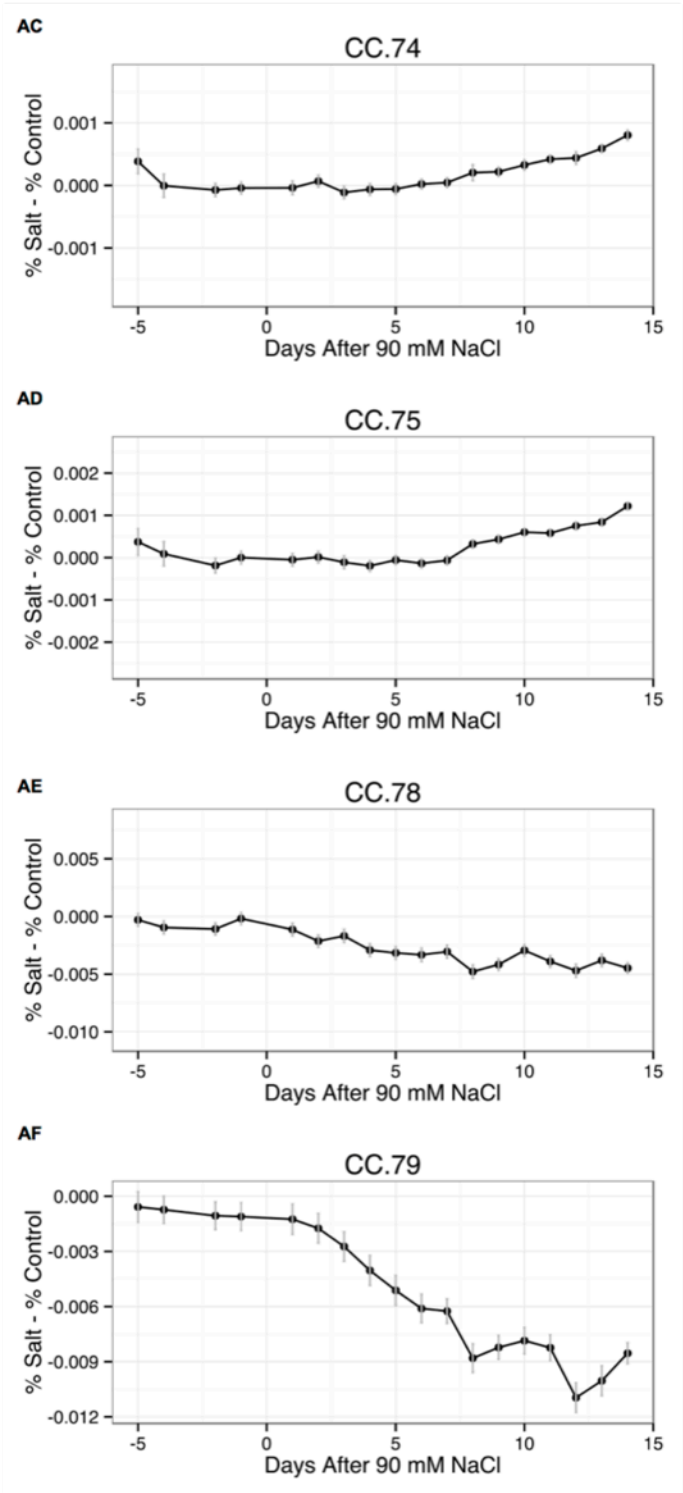

I
CC.22
0.006
0.003
0.000
-0.003
-0.006
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
J
CC.23
0.002
0.000
-0.002
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
K
CC.25
0.0010
0.0005
0.0000
-0.0005
-0.0010
-0.0015
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
L
CC.27
0e+00
-1e-04
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl


M
CC.32
% Salt - % Control
0.004
0.000
-0.004
-5
0
5
10
15
Days After 90 mM NaCl
N
CC.33
% Salt - % Control
0.0025
0.0000
-0.0025
-5
0
5
10
15
Days After 90 mM NaCl
O
CC.34
% Salt - % Control
0.003
0.002
0.001
0.000
-0.001
-0.002
-0.003
-5
0
5
10
15
Days After 90 mM NaCl
P
CC.37
% Salt - % Control
1e-04
0e+00
-1e-04
-2e-04
-3e-04
-4e-04
-5e-04
-5
0
5
10
15
Days After 90 mM NaCl


Q
CC.42
0.0050
0.0025
0.0000
-0.0025
-0.0050
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
R
CC.43
0.003
0.000
-0.003
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
S
CC.44
0.0050
0.0025
0.0000
-0.0025
-0.0050
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl
T
CC.5
0e+00
-1e-04
% Salt - % Control
-5
0
5
10
15
Days After 90 mM NaCl


u
CC.53
% Salt - % Control
0.002
0.001
0.000
-0.001
-0.002
-5
0
5
10
15
Days After 90 mM NaCl
v
CC.54
% Salt - % Control
0.004
0.002
0.000
-0.002
-0.004
-5
0
5
10
15
Days After 90 mM NaCl
w
CC.55
% Salt - % Control
0.004
0.002
0.000
-0.002
-0.004
-5
0
5
10
15
Days After 90 mM NaCl
x
CC.6
% Salt - % Control
5e-05
0e+00
-5e-05
-5
0
5
10
15
Days After 90 mM NaCl


Y
CC.6
% Salt - % Control
1e-04
0e+00
-1e-04
-5
0
5
10
15
Days After 90 mM NaCl
Z
CC.64
% Salt - % Control
0.003
0.002
0.001
0.000
-0.001
-0.002
-0.003
-5
0
5
10
15
Days After 90 mM NaCl
AA
CC.65
% Salt - % Control
0.004
0.002
0.000
-0.002
-0.004
-5
0
5
10
15
Days After 90 mM NaCl
AB
CC.68
% Salt - % Control
0.000
-0.001
-0.002
-0.003
-5
0
5
10
15
Days After 90 mM NaCl


AC
CC.74
% Salt - % Control
0.001
0.000
-0.001
-5
0
5
10
15
Days After 90 mM NaCl
AD
CC.75
% Salt - % Control
0.002
0.001
0.000
-0.001
-0.002
-5
0
5
10
15
Days After 90 mM NaCl
AE
CC.78
% Salt - % Control
0.005
0.000
-0.005
-0.010
-5
0
5
10
15
Days After 90 mM NaCl
AF
CC.79
% Salt - % Control
0.000
-0.003
-0.006
-0.009
-0.012
-5
0
5
10
15
Days After 90 mM NaCl

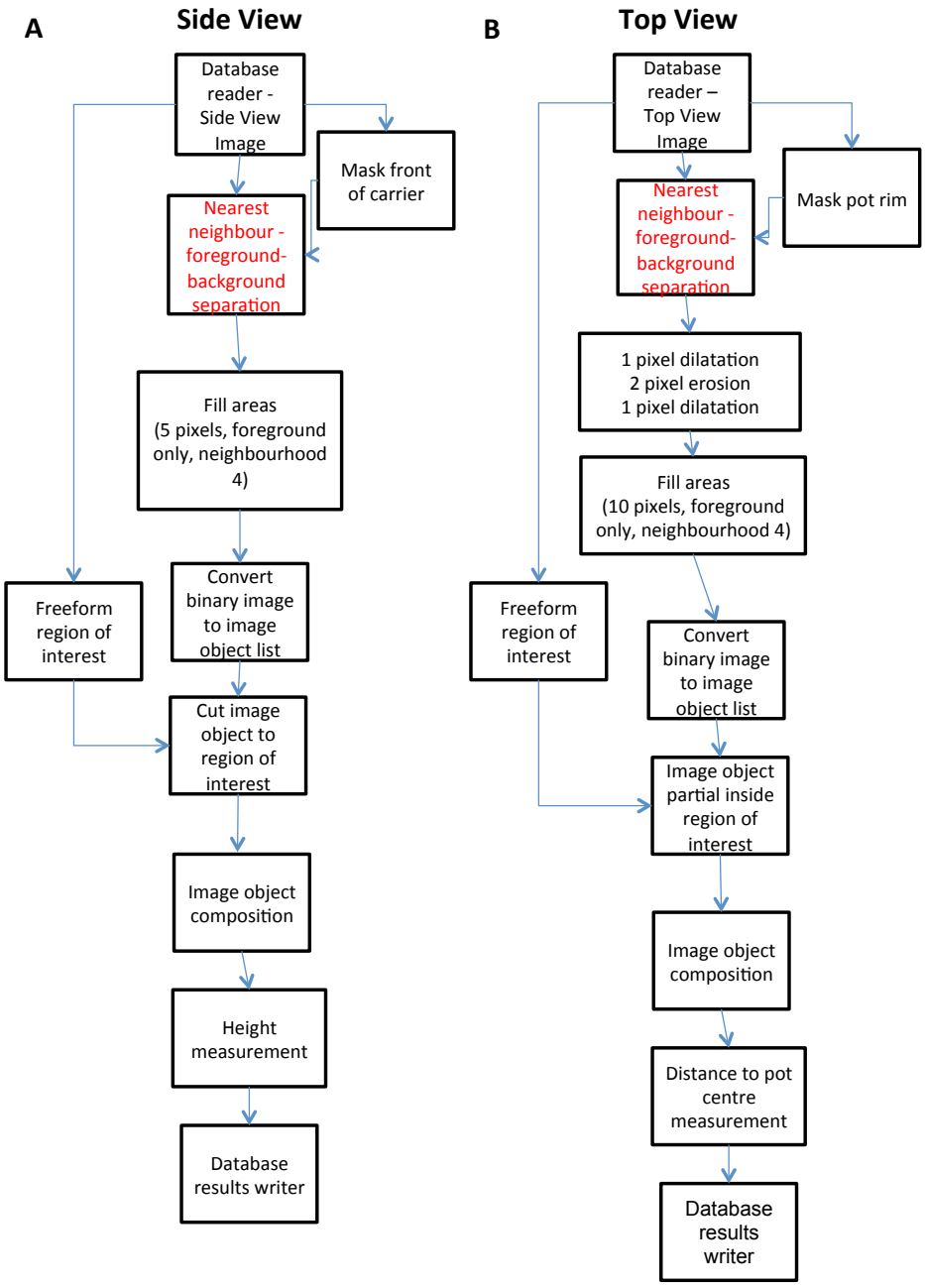


Figure S21. LemnaGrid pipeline used to process RGB side (A) and top (B) view images.

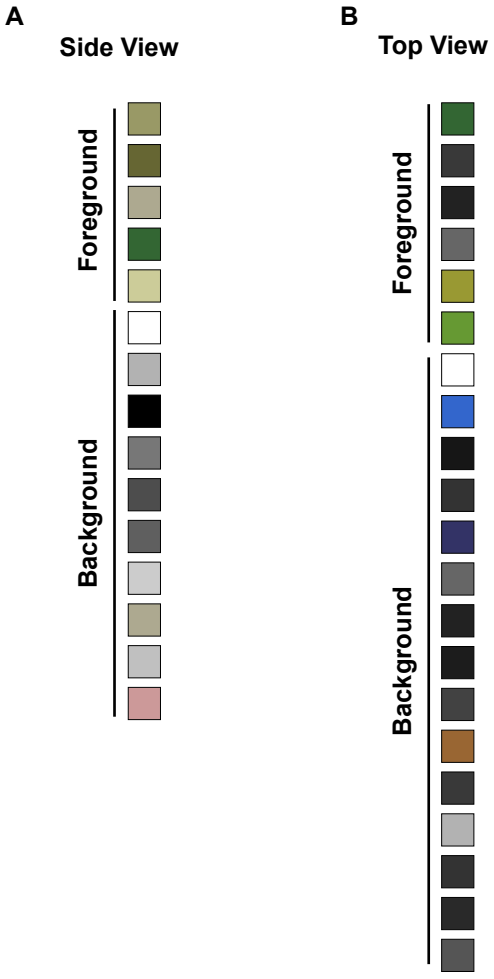


Figure S22. Color classes used to define foreground and background pixels from RGB side view (A) and top view (B) images for the LemnaGrid Nearest Neighbor foreground-background colour separation method.

S1 Table: Seven digital traits used to describe plant growth responses.

Abbreviation	Image Type	Processing Platform	View	Definition
MinR SV	RGB (Visable)	LemnaGrid	SV1 + SV2	The area of the minimum-enclosing rectangle enclosing the object.
MinC SV	RGB (Visable)	LemnaGrid	SV1 + SV2	The diameter of the minimum enclosing circle enclosing the object.
ConC SV	RGB (Visable)	LemnaGrid	SV1 + SV2	Circumference of the convex hull for the object.
ConA SV	RGB (Visable)	LemnaGrid	SV1 + SV2	Area of the convex hull for the object.
Ci SV	RGB (Visable)	LemnaGrid	SV1 + SV2	The circumference of the minimum circle that encloses the plant.
CL SV	RGB (Visable)	LemnaGrid	SV1 + SV2	The greatest distance between any of the points within the convex hull.
PSA	RGB (Visable)	LemnaGrid	SV1 + SV2 + TV	Projected Surface Area. Sum of plant pixels from each view.

The convex hull is the smallest shape that encloses a set of points *X*, so that any line drawn between points of *X* is enclosed within the shape.

Table S2. Accuracy of digital traits for describing biomass and shoot area. Pearson correlation analysis of seven growth-related digital metrics and three manual measurements. Asterisks indicate statistical significance ($p < 0.001$).

	MinR SV	MinC SV	ConC SV	ConA SV	Ci SV	CL SV	PSA	Dry Wt	Fresh Wt	Plant Area
MinR SV	-0.04	-0.1	0.35	0.38	0.34	0.25	0.96*	0.96*	0.96*	1
MinC SV	-0.11	-0.16	0.43*	0.47*	0.36	0.32	0.95*	0.97*	1	
ConC SV	-0.09	-0.13	0.4*	0.43*	0.35	0.3	0.95*	1		
ConA SV	-0.04	-0.11	0.43*	0.46*	0.38*	0.32	1			
Ci SV	-0.44*	-0.36	0.95*	0.89*	0.41*	1				
CL SV	-0.41*	-0.46*	0.44*	0.4*	1					
PSA	-0.5*	-0.47*	0.97*	1						
Dry Wt	-0.46*	-0.42*	1							
Fresh Wt	0.96*	1								
Plant Area	1									

Table S3. Candidate genes underlying significant SNPs. All genes within 200kb of a significant SNP were considered as possible candidates. A significance threshold of $p < 10^{-8}$ was used for logistic growth-response association analysis, and $p < 10^{-7}$ for GWA of salinity-induced fluorescence responses. SNP positions are based on MSU build 7.

Trait(s)	Chr.	SNP Position	SNP ID	p-value	Rice Gene	Rice Annotation	Arabidopsis Annotation		Reference(s)
Growth	1	7447351	dd1001171	1.89E-11	LOC_Os01g13460	Helix-loop-helix DNA-binding domain containing protein	AT2G46510	AIB	(Li et al., 2007)
Growth	2	32258316	id2014363	5.79E-10	LOC_Os02g52780	bZIP transcription factor	AT3G19290	ABF4	(Choi et al., 2005)
Growth	3	11079914	id3005794	2.25E-10	LOC_Os03g19720	EF hand family protein	AT5G49480	AtCP1	(Jang et al., 1998)
Growth	3	16293444	id3008139	1.11E-16	LOC_Os03g28120	Potassium channel protein	AT2G26650	AKT1	(Hirsch et al., 1998)
Growth	6	27647390	id6015813	2.81E-09	LOC_Os06g45940	Potassium transporter	AT2G30070	AtKT1	-
Growth	7	26461039	id7005111	9.45E-09	LOC_Os07g44290	CAMK-KIN1/SNF1/Nim1-like.29	AT3G17510	CIPK1	(D'Angelo et al., 2006)
Growth	9	879018	ud9000038	7.15E-09	LOC_Os09g02214	Citrate transporter protein	AT3G19490	AtNHD1	(Müller et al., 2014)
Growth	9	16562832	id9005164	2.05E-11	LOC_Os09g27580	Potassium transporter	AT2G30070	AtKT1	-
Growth	12	25277179	id12009128	1.17E-10	LOC_Os12g40830	PFKB kinase family	AT5G37850	AtSOS4	(Shi et al., 2002)
Growth	1	25003341	id1014913	3.18E-11	LOC_Os01g43410	CAMK-like.9	AT4G23650	CDPK6	(Latz et al., 2013)
FLUO (cc.41)	1	27738377	id1016079	2.57E-08	LOC_Os01g48680	Two pore calcium channel protein 1	AT4G03560	AtCCH1	(Choi et al., 2014)
FLUO (cc.41, cc.52)	1	29950678	id1017845	4.11E-09	LOC_Os01g52070	Potassium channel AKT1	AT4G32650	AtKC1	(Geiger et al., 2009)
FLUO (cc.31, cc.41, cc.52)	1	30454089	id1018311	6.52E-08	LOC_Os01g53294	Respiratory burst oxidase protein B	AT1G64060	RBOH F	(Ma et al., 2012)

Table S4. LemnaGrid devices used for processing RGB images and corresponding ImageHarvest/OpenCV functions.

LemnaGrid Device	LemnaGrid Parameters	Definition	ImageHarvest Function(s)*
Nearest Neighbor foreground-background colour separation	5 foreground colors, 10 background colors (see Fig. S21 for colors)	A set of colours are selected in that represent background and foreground objects. Each pixel in the image is assigned to background or foreground class using nearest neighbour method.	knn(self, k labels) self: class containing image k: number of forground colors classes labels: List of BGR (blue, green, red) ranges that define the plant/foreground object
Fill Holes	Five pixels (foreground only); neighborhood 4	A hole, which is defined as a background area that is completely surrounded by the foreground object, is filled using the user-defined parameters and is added to the foreground object.	cv2.findContours(img,cv2.RETR_CCOMP,cv2.CHAIN_APPROX_SIMPLE) img: An 8-bit single-channel image cv2.RETR_CCOMP: Contour retrieval mode (see OpenCV documentation for complete description) cv2.CHAIN_APPROX_SIMPLE: Contour approximation method (see OpenCV documentation for complete description) cv2.drawContours(img,contour) img: An 8-bit single-channel image
Morphological Filter	1 pixel dilation, 2 pixel erosion, 1 pixel dilation	This device performs a defined series of pixel size reduction steps on a binary image to eliminate small noise from the foreground object.	cv.Erode(src, dst, element=None, iterations=1) cv.Dilate(src, dst, element=None, iterations=1) src: source image dst: output image of the same size as src element: a structuring element used for erosion/dilation

File S1. Pearson correlation between PSA and 32 fluorescence color classes in control and saline conditions. Correlation analysis was performed at each time point individually.

File S2. Pearson correlation of salt-induced responses for PSA and fluorescence color classes. Correlation analysis was performed for salinity responses (i.e. PSA Salt / PSA Control or % Area Salt - % Area Control) at each time point.

File S3. Genes located within 200 kB of significant SNPs associated with salinity-induced growth responses identified with longitudinal GWA analysis ($p < 10^{-8}$).

File S4. Genes located within 200 kB of significant SNPs associated with salinity-induced growth responses identified with conventional mixed model GWA analysis ($p < 10^{-4}$).

File S5. Genes located within 200 kB of significant SNPs associated with salinity-induced fluorescence responses identified with conventional mixed model GWA analysis ($p < 10^{-7}$).

File S6. ANOVA results comparing RGB color classes between treatments. To identify color classes from RGB images displaying differences between treatments across all 373 accessions, a one-way blocked ANOVA was performed at each individual time-point, where treatment was considered as a fixed effect and accession was as a block. Color classes were that showed significant differences between treatments after three of more days of 90 mM NaCl were considered to be salinity-responsive ($p < 0.00056$).

File S7. Pearson correlation analysis of RGB color classes and four FLUO color classes. Pearson correlation analysis was performed using the 32 salinity responsive RGB color classes and FLUO color classes 21, 31, 41 and 52.

CHAPTER 6

TRANSCRIPTIONAL AND REGULATORY VARIATION IN SHOOT TISSUE UNDER SHORT-TERM SALINITY STRESS

6.1 Introduction

Salt tolerance can be simply defined as the ability to maintain growth or yield under saline conditions, or in other words, to lessen the effects of NaCl on agronomic performance. Growth can be impeded under saline conditions through several different processes. Shortly after the addition of NaCl to the soil solution, the high concentration of NaCl in the external media impedes water uptake, and reduces cell expansion. This osmotic effect of on shoot growth and carbon sequestration is evidenced by the nearly immediate reduction in stomatal conductance (Munns and Tester, 2008). The osmotic phase of salinity stress is experienced before Na^+ accumulates to toxic levels in the cell. The similar physiochemical properties of Na^+ and K^+ facilitates competition between the two ions, and results in the displacement of K^+ from sites (Zhu, 2007). Sodium toxicity occurs days or weeks after the onset of salinity, once Na^+ concentration surpasses a threshold and impedes cellular processes.

Plants have developed various mechanisms to adapt to the osmotic and ionic phases of salt stress. Compatible solutes such as proline, betaine, trehalose, or sucrose may be produced during the osmotic phase of the salt response to lower cell water potential and restore water influx to the cell. This process, known as osmotic adjustment (OA), has been documented in rice in response to water deficits, however the relevance of OA in restoring water relations during salinity stress is currently unknown (Lilley and Ludlow, 1996; Lilley et al., 1996; Babu et al., 1999). The mechanisms to tolerate the ionic phase of salt stress can be grouped into two types of

processes: Na^+ exclusion and tissue tolerance (Munns and Tester, 2008). The goal of Na^+ ionic tolerance mechanisms is to limit Na^+ toxicity in photosynthetically active tissues. For Na^+ exclusion mechanisms, this may be achieved by limiting the amount of Na^+ that enters the transpiration stream (limit Na^+ loading to xylem), reducing the amount of Na^+ that is unloaded from the xylem, or by storing Na^+ in older or less photosythetically active tissue (i.e. leaf sheaths) (Munns and Tester, 2008). Tissue tolerance may be achieved by sequestering Na^+ in the cell vacuole, production of compatible solutes, or increased metabolism of reactive oxygen species (Munns et al., 2006). The adaptive mechanisms underlying salt tolerance may not be mutually exclusive. These adaptive mechanisms may be dependent on the duration of stress, concentration of NaCl, or other environmental factors.

Considering the different strategies used to adapt to saline conditions, salt tolerance is genetically complex and involves many different genes (Horie and Schroeder, 2004; Munns and Tester, 2008; Horie et al., 2009; Roy et al., 2014). The genetic complexity of salt tolerance has hindered the development of varieties through traditional breeding approaches that have focused on growth or yield that can maintain growth or yield under saline conditions. Considerable progress has been made in understanding the genetic basis of physiological traits, such as Na^+ exclusion, that contribute to salt tolerance. However, the genetic basis of tissue tolerance and osmotic tolerance is limited. While improved salt tolerance can be achieved by incorporating loci regulating these traits into rice varieties, key regulators of additional adaptive strategies are necessary.

Over the past decade genome-wide association studies (GWAS) have proven to be a powerful approach to identify variants associated with agronomically important traits. While this approach has provided an effective means to study the genetic architecture of many traits, the

identification of causal genes and functional DNA elements remains a major challenge in uncovering the genetic mechanisms that contribute to phenotypic variation.

Following the central dogma of molecular biology, causal variants should change the production of RNA and/or protein, which will affect the system as a whole and thus lead to perturbation in the phenotype. The utilization of intermediate phenotypes, such as gene expression, metabolites, or protein levels, is a natural step forward in the dissecting the genetic mechanisms underlying phenotypic variation. Gene expression measurements collected from a diverse population of individuals can be used in the standard statistical framework used for association mapping to identify genetic variants that may affect gene expression (eQTL) (Flutre et al., 2013; Peters et al., 2016; Zhu et al., 2016). The colocalization of eQTL and phenotypic QTL (pQTL) can greatly accelerate the selection of potential causal variants and/or genes that can be functionally validated through transgenic approaches.

Several methods have been developed to integrate transcriptome data and GWAS summary statistics in an effort to understand the genetic basis of complex traits. For instance, the method developed by Giambartolomei *et al* utilizes a Bayesian approach to identify eQTL SNPs that colocalize with phenotypic QTL SNPs (Giambartolomei et al., 2014). With this approach the probability of both traits being regulated by a single common causal variant can be formally tested. The advantages of this approach is that it can be run using only summary statistics from GWAS studies, and that it accounts for local LD at the region of interest. However, when the causal SNP only accounts for a small portion of the phenotypic variance this approach may be underpowered (Giambartolomei et al., 2014).

Rather than identifying eQTL that colocalize with phenotypic QTL, the approach developed by Gusev *et al* tests for associations between gene expression and the trait of interest

(Gusev et al., 2016). Briefly, the method utilizes two populations: one smaller population where the individuals are genotyped and also have gene expression measurements, and the second larger population that is genotyped and phenotyped. For each gene, the local *cis*-genetic effects for gene expression are predicted in the smaller population, and these weights are correlated with the SNP effects on the phenotype of interest determined from the larger panel. This approach facilitates the identification of causal genes as well as those that are downstream of the causal gene, but may participate in the development of the phenotype.

In this study we leveraged RNA sequencing data from 91 diverse rice accessions from the Rice Diversity Panel 1 to understand the effects of short-term salt stress on gene expression in shoot tissue, and identify genes associated with complex salt-tolerance related traits (Zhao et al., 2011a). This information can be used to better understand how plants cope with short-term salt stress, and facilitate the identification of genes underlying important traits.

6.2 Results and discussion

The primary goals of this study were to (1) assess natural variation in molecular responses to salt stress and (2) identify key regulators underlying these complex responses. To this end, RNA sequencing was performed on shoot tissue collected from ninety-one diverse accessions of RDP1 shorty after the onset of mild salt stress (6 dS m^{-1}). The accessions were selected to represent four subpopulations of *Oryza sativa*: *Aus*, *Indica*, *Tropical Japonica* and *Temperate Japonica*.

To identify the major sources of variation in the transcriptome dataset, PCA analysis was conducted using variance stabilized read counts from all 361 samples. Here, genes with 0 variance were excluded from the PCA analysis, leaving 51,944 genes for PCA analysis. The samples clustered into two distinct groups along the first principle component, which accounted

for 7.6% of phenotypic variation (Figure 1). Samples belonging to the same cluster belonged largely to the same subspecies (*Indica* or *Japonica*), rather than treatment groups. This suggests that the subspecies classification, rather than treatments, has a much larger effect on the overall variation in transcriptomes in the dataset.

6.2.1 Differential expression analysis

To identify genes that may be affected by the mild salinity treatment, a mixed model was fit in "limma" (Ritchie et al., 2015). To account for the large effect of subspecies on overall transcriptional variation, the model included both subspecies and treatment as fixed effects, an interaction term was included to examine the differential effects of subspecies on transcriptional responses to salinity, and accession was included as a random effect. In agreement with the results of the PCA analysis, far more genes were found to be affected by subspecies (18,211 genes) compared to treatment (100 genes). These results further support the results from the PCA analysis and suggest that variation in the transcriptomes is driven largely by population structure, rather than the salinity treatment. No genes were identified that were affected by the interaction between treatment and subspecies.

To identify biological processes that may be affected by salinity treatment, gene set enrichment analysis was conducted on the list of 100 differentially expressed genes. Several genes that have been shown to be components of the transcriptional responses to dehydration, and are responsive to excess NaCl in multiple species were differentially expressed between treatments. Three genes encoding CBF/DREB homologues, which are known to be important for regulating the responses to cold and high-salt stress in rice, were expressed at significantly higher levels in shoot tissue of salt stressed plants relative to control plants (Ito et al., 2006). Moreover, the expression of *CIPK1*, which is known to regulate ABA signaling responses in

Arabidopsis during osmotic stress, was induced in salt stressed plants (D'Angelo et al., 2006). The induction of these osmotic stress responsive genes suggests that cell water relations may be disrupted shortly after the onset of mild salinity stress.

Genes associated with ion transport, cation ion transport and transition metal ion homeostasis were significantly enriched among the differentially expressed genes ($p < 0.05$, Table I). Several genes that have previously been shown to regulate iron uptake and/or transport showed significant differences in expression between treatments (Table I). To further characterize this gene set in response to iron deficiency, the expression of each gene was examined using a publically available dataset of rice root and shoot tissue under iron deficient conditions (GSE17245). A total of 880 genes were found to be differentially expressed in response to iron deficient conditions in shoot tissue, while a total of 1352 genes were found to be differentially expressed between treatments in root tissue (BH adjusted *p*-value 0.001; $|LFC| \geq 1.5$). Interestingly, 35 (44 probe sets) of the 100 genes found to be differentially expressed in response to salinity treatment in shoot tissue were also found to be differentially expressed in response to iron deficiency in shoots (Table II).

OsIRO3, which is homologous to the *Arabidopsis* gene *PYE*, showed higher expression in control plants relative to salt stress plants. Moreover, homologous of several genes that have been shown to be coexpressed with *PYE* in *Arabidopsis* showed significantly higher expression in control plants compared to salt stressed plants (*NRAMP*, *IRO3*, and *OPT3*) (Stein and Waters, 2011). On the other hand, genes that function in the storage of iron during sufficient conditions were expressed at higher levels in shoot tissue of salt stressed plants relative to controls. These results suggest that iron homeostasis may be perturbed shortly after the onset of salt stress.

6.2.2 eQTL

To identify loci underlying natural variation in gene expression, *cis*-*e*QTLs were jointly modeled in across treatments using 220,083 SNPs, and 24,293 genes. This approach has been shown to improve power to detect *cis*-eQTL in multiple tissues in humans compared to conventional approaches that utilized linear models to identify SNP-gene associations in single conditions (Flutre et al., 2013). Moreover, the probability of a gene-SNP association being detected in within or across conditions can be formally determined.

A total of 7725 genes we found to be associated with one or more SNP. However, no SNP-associations were identified that were specific to one condition, the 7725 genes were found to be present in both salt and control conditions. The lack of condition-specific *cis*-*e*QTL may be due to the relatively mild salt treatment imposed during the experiment, which is supported by the low number of differentially expressed genes and the cluster of samples according to population/subspecies rather than treatment.

6.2.3 A transcriptome-wide association study identifies genes potentially underlying complex salt-tolerance related traits

To prioritize the selection of candidate genes for functional analysis, we utilized a method developed by Gusev *et al* named TWAS (transcriptome-wide association study; Gusev et al., 2016). Briefly, the method utilizes transcriptome data from a small reference population, in this case 91 accessions of RDP1, to predict the *cis* genetic effects of gene expression, and GWAS summary statistics derived from a larger population, here ~380 accessions of RDP1. Using the transcriptome data from the 91 accessions of RDP1, the *cis* genetic effects of 4432 genes could be predicted. The heritability estimates of the 4432 genes ranged from 0.045 to 0.743, with a

total of 807 genes displaying moderate heritability ($h^2 \geq 0.4$) (Figure 2). Of these 807 genes, 82 displayed high heritability ($h^2 \geq 0.6$).

GWAS summary statistics were generated for eight salt-tolerance phenotypes using a linear mixed model that used the first four principal components and kinship matrix to correct for population structure and relatedness among lines using GAPIT (Lipka et al., 2012). The eight salt-tolerance phenotypes collected were root and shoot growth responses, root Na^+, root Na^+:K^+, root K^+, shoot Na^+, shoot N^+:K^+, and shoot K^+. Using the TWAS approach, a total of 224 non-redundant genes were found to be significantly associated with one or more of the eight phenotypes (Table III; $p \leq 2.26 \times 10^{-5}$).

To identify biological processes that may be over-represented within these gene sets, gene set enrichment analysis was conducted for each trait. Genes associated with ion transport were over-represented for shoot Na^+:K^+, and root Na^+:K^+ and K^+ (Table III). A gene annotated as a sodium:calcium exchanger (LOC_Os02g14980) was found to be positively associated with shoot Na^+:K^+, indicating that accessions with high shoot Na^+:K^+ also exhibit higher expression of this Na^+:Ca^+ exchanger. For root Na^+:K^+ and root K^+ a gene encoding a potassium channel (LOC_Os06g14310), similar to the *Arabidopsis* gene *KC1*, was found to be significantly associated with both traits. *KC1* was found to be positively associated with root Na^+:K^+, indicating that lines with high expression of *KC1* in shoot tissue exhibit higher Na^+:K^+ in the roots. For root K^+, *KC1* showed a negative association, meaning that plants with high expression of *KC1* in shoot tissue also show low K^+ content in roots. In *Arabidopsis*, *KC1* interacts with the potassium channels *AKT* and *KAT*, and inhibit K^+ transport (Reintanz et al., 2002; Wang et al., 2010). While this gene has yet to be characterized in rice, the role of this gene in *Arabidopsis* is in agreement with the direction of TWAS associations.

6.3 Materials and Methods

6.3.1 Growth conditions for GWAS studies

A detailed description of growth conditions and trait measurements are provided in chapter 2. Briefly, the study included 383 of the 421 original RDP1 accessions, as well as seven check varieties (Zhao et al., 2011b; Famoso et al., 2011; Eizenga et al., 2014). A total of 365 accessions from RDP1 were genotyped using 700,000 SNPs (McCouch et al., 2016).

The experiment was conducted between July to Sep 2013 in a controlled green house at Lincoln, NE. Seedlings were transplanted into the pots filled with Turface (Profile Products, LLC) and were grown in tap water for four days after transplanting. For the remainder of the experiment the plants were supplemented with half strength Yoshida solution (pH 5.8) (Yoshida et al., 1976). NaCl was mixed with $CaCl_2$ in a 6:1 molar ratio and was added after 10 d of seedling growth. The stress treatment was started at 2.5 $dS \cdot m^{-1}$ which increased gradually up to 9.5 $dS \cdot m^{-1}$ in 4 steps over a period of four days (~2 $dS \cdot m^{-1}$ or 20 mM NaCl per day) to avoid any osmotic shock to the plants. The stress treatment was stabilized at 9.5 $dS \cdot m^{-1}$ for next two weeks. Root and shoot samples were collected separately and rinsed 3 times in tap water and once in deionized water to remove excess NaCl at the completion of the experiment (14 days of 9.5 $dS \cdot m^{-1}$; 28 days after transplant). The samples were oven dried at 60 °C for one week prior to measuring root and shoot biomass. Shoot and roots from two plants were taken for biomass measurement.

Dried shoot samples were ground and 200 – 300 g of total material was digested with 0.1N Nitric acid (Fisher Scientific) at 70 °C for 8 hrs. Root samples were weighed and digested without any grinding. Samples were diluted and cation (Na^+ and K^+) concentrations in the plant

extract were determined with appropriate standard by dual Flame photometry (Cole Parmer, USA).

Phenotypic data was combined across periods and a linear model was fit to calculate adjusted means for individual accession using the PROC GLM procedure of the Statistical Analysis System (SAS Institute, Inc.). The linear model included period (i.e., June-July or Aug-Sept), replication nested within period, tub nested within replication, accession, and accession-by-period interaction.

The linear mixed model $y= X\beta + C\gamma + Zu + e$ was used to detect SNPs associated with each trait. The model was implemented using the efficient mixed-model association (EMMA) algorithm of Kang *et al* (Kang et al., 2008). VanRaden's method was used to generate the kinship matrix (VanRaden, 2008). The GWAS analysis was conducted using GAPIT (Lipka et al., 2012)

6.3.2 Growth conditions for RNA sequencing study

For gene expression analysis plants from 91 accessions were grown in a controlled environment growth chamber. The temperatures were maintained at 28 °C and 25 °C in day and night respectively and 60% relative humidity. Lighting was maintained at 800 micromoles m^{-2} s^{-1} using high-pressure sodium lights (Phillips). Seeds preparation and salt treatment were performed as described above. Eight day (four days after transplant) old rice seedlings were subjected to 6 $dS \cdot m^{-1}$ for a period of 24h. Salinity stress was increased to 6 $dS \cdot m^{-1}$ gradually in two 3 $dS \cdot m^{-1}$ intervals over a period of 24h. After 24h of 6 $dS \cdot m^{-1}$, aerial parts of the seedlings were excised from the roots and frozen immediately in liquid nitrogen. The samples were ground with Tissuelyser II (Invitrogen) and total RNA was isolated with RNAeasy isolation kit (Qiagen) according to manufacturer's instructions. On-column DNAse treatment was performed to

remove genomic DNA contamination (Qiagen). Sequencing was performed using Illumina HiSeq 2500. Sixteen RNA samples were combined in each lane. Two biological replicates were used for each accession.

6.3.3 RNA-seq mapping and analysis

Read mapping is identical to the methods described in Chapter 4. The Bioconductor packages 'limma' and 'EdgeR' were used to identify genes that were differentially expressed between treatments (Gentleman et al., 2004; Robinson and Oshlack, 2010; Ritchie et al., 2015). The following linear model was fit using limma $y = subspecies + treatment + treatment$ x *subspecies*. Genes with a significant treatment term were selected for further analysis. Benjamini and Hochberg's method was used to control the false discovery rate, and genes with an adjusted p-value ≤ 0.05 were considered differentially expressed (Benjamini and Hochberg, 1995).

6.3.4 Microarray processing and analysis

Publically available microarray data generated by Zheng *et al* (GSE17245) were obtained from Gene Expression Omnibus (GEO; Zheng et al., 2009). Arrays from plants grown under P deficiency were discarded. The CEL files were analyzed using the affy and limma packages in R (Gautier et al., 2004; Ritchie et al., 2015). The arrays were normalized using robust multi-array (RMA) approach. Contrasts were made between iron sufficient and iron deficient conditions within tissues. Correction for multiple testing was done using Benjamini and Hochberg's method (Benjamini and Hochberg, 1995). Genes were considered to be differentially expressed if BH-adjusted p-value < 0.001 and a log fold change of ± 1.5.

6.3.5 Cis-eQTL discovery

Preparation of RNA-seq data. Raw read counts for all 361 samples from both control and salt treated plants were loaded into DEseq2, and variance stabilizing transformation was performed to reduce the dependence of the variance of gene expression on the mean (Love et al., 2014). Genes with a read count of less than 10 in less than 10% of the 361 samples were removed from the analysis. To ensure a normal distribution the expression of each gene was normalized using quantile normalization with random breaking of ties.

Preparation of genotypic data. High Density Rice Array (HDRA) data was obtained from ricediversity.org. Genotypic information for RPD1 accessions was extracted and missing genotypes were imputed using Beagle with 20 iterations (Browning and Browning, 2016). Markers with a maf < 0.05 were excluded.

Cis-eQTL discovery with BMA-eQTL. Multi-trait Bayesian linear regression models were used to jointly test for eQTLs for three models (Flutre et al., 2013). The method developed by Flutre *et al* essentially detects SNP-gene associations within each condition using a generalized linear model, and uses summary statistics in a hierarchical model to estimate the eQTL sharing among treatments. In each treatment the following linear model was fit $y= X\beta + C\gamma + e$, where β is the SNP effect; γ is a vector of PCs effects to account for population structure; X is a vector of SNP genotypes; C is an incidence matrix relating γ to y which consists of the first four principal components (PCs). The approach is Bayesian in that it allows for explicit testing of specific hypotheses, or configurations. In this study the hypotheses were (1) the eQTL is active in only control conditions, (2) the eQTL is active in only salt conditions, and (3) the eQTL is active in both conditions. The probability of each configuration or hypothesis is determined using a hierarchical model that borrows information across genes. To compute gene-level null priors (π_o)

the EBF/QBF procedures developed by Wen *et al* was used (Wen, 2013). Gene expression data was normalized using a variance stabilizing transformation in DEseq2 and were quantile normalized to ensure a normal distribution (Love et al., 2014).

6.3.6 Transcriptome-wide association study (TWAS)

The TWAS method developed by Gusev *et al* was used to identify genes that are associated with salinity tolerance related traits (Gusev et al., 2016). The TWAS method requires two datasets: the first data set is a list of gene weights is essentially the *cis*-genetic contribution to gene expression, and GWAS summary statistics for the eight salinity tolerance related traits. For each gene, SNPs within the 100 kb of the transcription start site was extracted using plink with the MSUv7 annotations (Chang et al., 2015). The quantile normalized gene expression measurements were used for TWAS analysis. The gene weights were calculated using the function "FUSION.compute_weights" in the program Fusion (Gusev et al., 2016). Summary statistics obtained using GAPIT (described above) was used to test for associations between gene expression and salt tolerance related traits using function "FUSION.assoc_test" in Fusion. The Bonferroni correction was used to test for statistical significance at FDR 0.1, which corresponds to an actual *p*-value of 2.26×10^{-5}.

6.3.7 Gene set enrichment analysis (GSEA)

GSEA was conducted using the R packages "GOstats", "GSEAbase", and "GO.db" (Morgan et al., 2013). A threshold of 0.05 was used to identify significant terms associated with biological processes and molecular function categories.

6.4 References

Babu, R.C., M.S. Pathan, A. Blum, and H.T. Nguyen. 1999. Comparison of measurement methods of osmotic adjustment in rice cultivars. Crop Sci. 39(1): 150–158.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B: 289–300.

Browning, B.L., and S.R. Browning. 2016. Genotype imputation with millions of reference samples. Am. J. Hum. Genet. 98(1): 116–126.

Chang, C.C., C.C. Chow, L.C.A.M. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4(1): 7.

D'Angelo, C., S. Weinl, O. Batistic, G.K. Pandey, Y.H. Cheong, S. Schültke, V. Albrecht, B. Ehlert, B. Schulz, K. Harter, S. Luan, R. Bock, and J. Kudla. 2006. Alternative complex formation of the Ca2+-regulated protein kinase CIPK1 controls abscisic acid-dependent and independent stress responses in Arabidopsis. Plant J. 48(6): 857–872.

Eizenga, G.C., M. Ali, R.J. Bryant, K.M. Yeater, A.M. McClung, S.R. McCouch, and others. 2014. Registration of the rice diversity panel 1 for genomewide association studies. J. Plant Regist. 8(1): 109–116.

Famoso, A.N., K. Zhao, R.T. Clark, C.-W. Tung, M.H. Wright, C. Bustamante, L. V Kochian, and S.R. McCouch. 2011. Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7(8): e1002221.

Flutre, T., X. Wen, J. Pritchard, and M. Stephens. 2013. A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet 9(5): e1003486.

Gautier, L., L. Cope, B.M. Bolstad, and R.A. Irizarry. 2004. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20(3): 307–315.

Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, and others. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5(10): 1.

Giambartolomei, C., D. Vukcevic, E.E. Schadt, L. Franke, A.D. Hingorani, C. Wallace, and V. Plagnol. 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10(5): e1004383.

Gusev, A., A. Ko, H. Shi, G. Bhatia, W. Chung, B.W.J.H. Penninx, R. Jansen, E.J.C. De Geus, D.I. Boomsma, F.A. Wright, and others. 2016. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet.

Horie, T., F. Hauser, and J.I. Schroeder. 2009. HKT transporter-mediated salinity resistance mechanisms in Arabidopsis and monocot crop plants. Trends Plant Sci. 14(12): 660–668.

Horie, T., and J.I. Schroeder. 2004. Sodium transporters in plants. Diverse genes and

physiological functions. Plant Physiol. 136(1): 2457–2462.

Ito, Y., K. Katsura, K. Maruyama, T. Taji, M. Kobayashi, M. Seki, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2006. Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. Plant Cell Physiol. 47(1): 141–153.

Kang, H.M., N. a Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178(3): 1709–23.

Lilley, J.M., and M.M. Ludlow. 1996. Expression of osmotic adjustment and dehydration tolerance in diverse rice lines. F. Crop. Res. 48(2): 185–197.

Lilley, J.M., M.M. Ludlow, S.R. McCouch, and J.C. O'toole. 1996. Locating QTL for osmotic adjustment and dehydration tolerance in rice. J. Exp. Bot. 47(9): 1427–1436.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28(18): 2397–2399.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12): 550.

McCouch, S.R., M.H. Wright, C.-W. Tung, L.G. Maron, K.L. McNally, M. Fitzgerald, N. Singh, G. DeClerck, F. Agosto-Perez, P. Korniliev, and others. 2016. Open access resources for genome-wide association mapping in rice. Nat. Commun. 7.

Morgan, M., S. Falcon, R. Gentleman, M.B.P. Maintainer, I. AnnotationDbi, G. biocViews GeneExpression, and G.O. GraphAndNetwork. 2013. Package "GSEABase."

Munns, R., R.A. James, and A. Läuchli. 2006. Approaches to increasing the salt tolerance of wheat and other cereals. J. Exp. Bot. 57(5): 1025–1043.

Munns, R., and M. Tester. 2008. Mechanisms of salinity tolerance. Annu. Rev. Plant Biol. 59: 651–81.

Peters, J.E., P.A. Lyons, J.C. Lee, A.C. Richard, M.D. Fortune, P.J. Newcombe, S. Richardson, and K.G.C. Smith. 2016. Insight into genotype-phenotype associations through eQTL mapping in multiple cell types in health and immune-mediated disease. PLoS Genet 12(3): e1005908.

Reintanz, B., A. Szyroki, N. Ivashikina, P. Ache, M. Godde, D. Becker, K. Palme, and R. Hedrich. 2002. AtKC1, a silent Arabidopsis potassium channel $α$-subunit modulates root hair K+ influx. Proc. Natl. Acad. Sci. 99(6): 4079–4084.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.: gkv007.

Robinson, M.D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11(3): 1.

Roy, S.J., S. Negrão, and M. Tester. 2014. Salt resistant crop plants. Curr. Opin. Biotechnol. 26: 115–124.

Stein, R.J., and B.M. Waters. 2011. Use of natural variation reveals core genes in the transcriptome of iron-deficient Arabidopsis thaliana roots. J. Exp. Bot. 63(2): 1039–1055.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11): 4414–4423.

Wang, Y., L. He, H.-D. Li, J. Xu, and W.-H. Wu. 2010. Potassium channel $α$-subunit AtKC1 negatively regulates AKT1-mediated K+ uptake in Arabidopsis roots under low-K+ stress. Cell Res. 20(7): 826–837.

Wen, X. 2013. Robust Bayesian FDR Control Using Bayes Factors, with Applications to Multi-tissue eQTL Discovery. Stat. Biosci.: 1–22.

Yoshida, S., D.A. Forno, and J.H. Cock. 1976. Laboratory Manual for Physiological Studies of Rice.

Zhao, K., C.-W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011a. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2(1–4): 467.

Zhao, K., C.W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011b. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Commun 2: 467Available at http://www.ncbi.nlm.nih.gov/pubmed/21915109%5Cnhttp://www.nature.com/ncomms/journal/v2/n9/pdf/ncomms1467.pdf.

Zheng, L., F. Huang, R. Narsai, J. Wu, E. Giraud, F. He, L. Cheng, F. Wang, P. Wu, J. Whelan, and others. 2009. Physiological and transcriptome analysis of iron and phosphorus interaction in rice seedlings. Plant Physiol. 151(1): 262–274.

Zhu, J.K. 2007. Plant salt stress. Encyclopedia of Life Science.

Zhu, Z., F. Zhang, H. Hu, A. Bakshi, M.R. Robinson, J.E. Powell, G.W. Montgomery, M.E. Goddard, N.R. Wray, P.M. Visscher, and others. 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet.

6.5 Figures

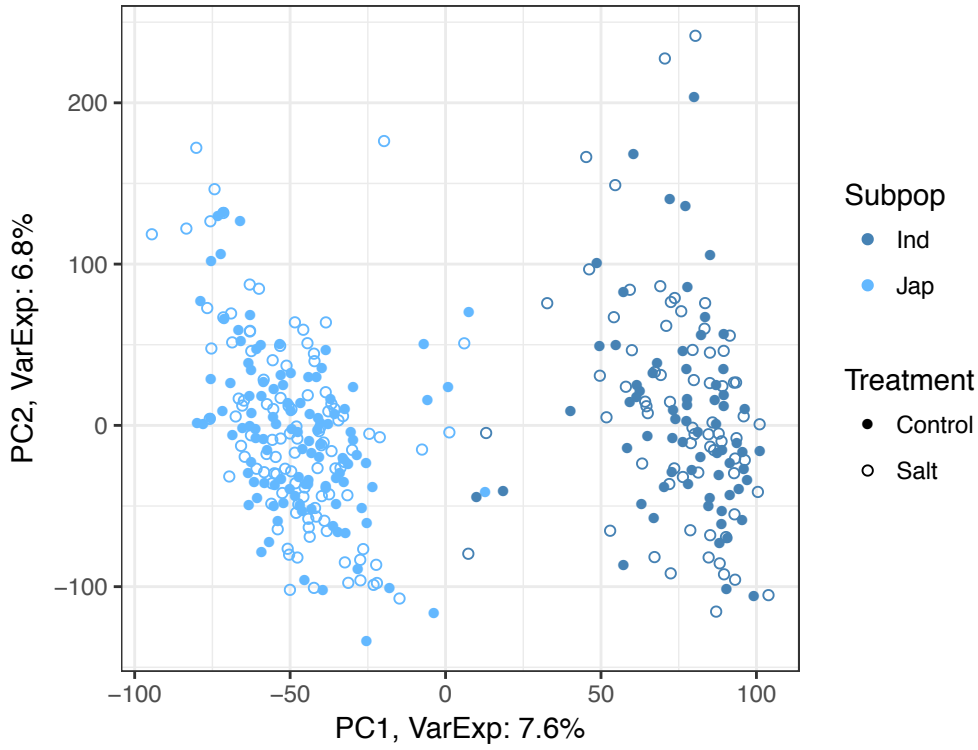


Figure 1. Principal component analysis of gene expression for 91 accessions in salt and control conditions. Open circles indicated salt treated samples while closed circles indicate control plants. Subspecies assignment was done using STRUCTURE with K=2.

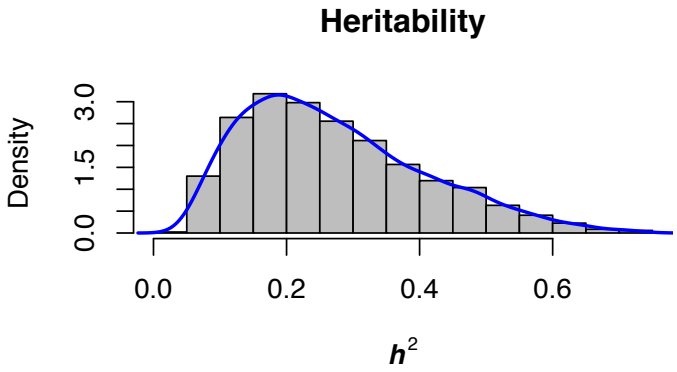


Figure 2. Histogram of heritability estimates for 4432 genes used for TWAS. Local *cis* heritability was determined for 4432 using the TWAS pipeline implemented in Fusion (Gusev et al., 2016).

6.6 Tables

Table I. Subset of significant GO terms identified through gene set enrichment analysis.

GSEA was conducted for genes that were differentially expressed between salt and control plants.

GO ID	P-value	Term	GO Category
GO:0030003	0.009	cellular cation homeostasis	BP
GO:0006873	0.009	cellular ion homeostasis	BP
GO:0006875	0.009	cellular metal ion homeostasis	BP
GO:0006879	0.009	cellular iron ion homeostasis	BP
GO:0006826	0.009	iron ion transport	BP
GO:0055065	0.009	metal ion homeostasis	BP
GO:0055072	0.009	iron ion homeostasis	BP
GO:0055076	0.009	transition metal ion homeostasis	BP
GO:0006812	0.016	cation transport	BP
GO:0055080	0.022	cation homeostasis	BP
GO:0048878	0.022	chemical homeostasis	BP
GO:0050801	0.022	ion homeostasis	BP
GO:0098771	0.022	inorganic ion homeostasis	BP
GO:0006811	0.023	ion transport	BP
GO:0000041	0.026	transition metal ion transport	BP
GO:0015992	0.050	proton transport	BP
GO:0006818	0.050	hydrogen transport	BP

Table II. Subset of genes differentially expressed in shoot tissue during salt treatment and iron deficiency.

Rice Locus	Probe ID	logFC (+Fe vs -Fe)	P-value	logFC (Salt vs Cont)	Rice Annotation	Arabidopsis Gene ID	Arabidopsis Gene Name
LOC_Os09g23300	Os.9660.1.S1_at	-6.48	2.96E-16	1.62	integral membrane protein, putative, expressed	AT2G01770.1	ATVIT1,VIT1
LOC_Os09g23300	OsAffx.6372.1.S1_s_at	-5.55	5.89E-15	1.62	integral membrane protein, putative, expressed	AT2G01770.1	ATVIT1,VIT1
LOC_Os11g01530	Os.12096.4.S1_s_at	-4.18	9.66E-13	1.1	ferritin-1, chloroplast precursor, putative, expressed	AT2G40300.1	ATFER4,FER4
LOC_Os12g01530	Os.12096.3.S1_a_at	-4.1	1.36E-15	1.06	ferritin-1, chloroplast precursor, putative, expressed	AT2G40300.1	ATFER4,FER4
LOC_Os05g34830	Os.37548.1.S1_at	-3.63	1.25E-14	0.53	No apical meristem protein, putative, expressed	AT1G01720.1	ANAC002,ATAF1
LOC_Os11g01530	Os.12096.3.S1_x_at	-3.5	2.23E-15	1.1	ferritin-1, chloroplast precursor, putative, expressed	AT2G40300.1	ATFER4,FER4
LOC_Os09g35010	Os.5816.1.S1_at	-2.9	5.62E-13	1.14	dehydration-responsive element-binding protein, putative, expressed	AT4G25480.1	ATCBF3,CBF3,DREB1A
LOC_Os03g12510	Os.12163.1.S1_at	-2.7	1.18E-11	-1.23	non-symbiotic hemoglobin 2, putative, expressed	AT2G16060.1	AHB1,ARATH GLB1,ATGLB1,GLB1,HB1,NSHB1
LOC_Os03g12510	Os.12191.1.S1_s_at	-2.48	9.40E-13	-1.23	non-symbiotic hemoglobin 2, putative, expressed	AT2G16060.1	AHB1,ARATH GLB1,ATGLB1,GLB1,HB1,NSHB1
LOC_Os07g43530	Os.34549.1.S1_at	-2.43	8.23E-10	0.6	helix-loop-helix DNA-binding domain containing protein, expressed	AT3G47640.1	PYE
LOC_Os03g58170	Os.11421.2.S1_x_at	-2.35	1.95E-12	0.27	stem-specific protein TSJT1, putative, expressed	AT3G22850.1	-
LOC_Os03g58170	Os.11421.1.S1_at	-2.14	1.69E-11	0.27	stem-specific protein TSJT1, putative, expressed	AT3G22850.1	-
LOC_Os11g38870	Os.27243.1.S1_at	-1.88	9.16E-10	0.44	helix-loop-helix DNA-binding domain containing protein, expressed	AT3G47640.1	PYE
LOC_Os06g35960	Os.10570.1.S1_at	-1.71	2.53E-09	0.37	HSF-type DNA-binding domain containing protein, expressed	AT3G24520.1	AT-HSFC1,HSFC1
LOC_Os04g48930	Os.49111.1.S1_at	1.66	1.08E-08	-1.53	ferric-chelate reductase, putative, expressed	AT1G01590.1	ATFRO1,FRO1
LOC_Os03g54000	Os.4978.1.S1_at	3.79	3.61E-13	-0.39	oligopeptide transporter, putative, expressed	AT4G16370.1	ATOPT3,OPT3
LOC_Os03g54000	Os.4978.1.S2_at	3.98	5.84E-14	-0.39	oligopeptide transporter, putative, expressed	AT4G16370.1	ATOPT3,OPT3
LOC_Os03g26210	Os.9216.1.S1_at	5.09	1.82E-13	-0.94	helix-loop-helix DNA-binding domain containing protein, expressed	AT3G47640.1	PYE
LOC_Os07g15460	Os.409.1.S1_at	6.07	3.78E-16	-1.31	metal transporter Nramp6, putative, expressed	AT1G80830.1	ATNRAMP1,NRAMP1,PMIT1

Table III. Summary of TWAS results for eight salinity tolerance related traits.

Trait	No of TWAS Genes	GO ID	GSEA P-value	Rice Gene	Rice Annotation	Arabidosis ID	Arabidopsis Gene Name	Arabidopsis Description	TWAS-Z	TWAS P-value
Shoot Na:K	39	calcium ion binding (GO:0005509)	0.003	LOC_Os02g14980	calcium-binding EF hand family protein, putative, expressed	AT1G53210	-	sodium/calcium exchanger family protein	4.34	1.41E-05
		ion binding (GO:0043167)	0.028							
		anion binding (GO:0043168)	0.029							
Shoot Na	45	ubiquitin-protein transferase activity (GO:0004842)	0.026	-	-	-	-	-	-	-
		ubiquitin-like protein transferase activity (GO:0019787)	0.040							
Shoot K	32	transcription factor activity, transcription factor binding (GO:0000989)	0.005	-	-	-	-	-	-	-
		adenyl ribonucleotide binding (GO:0032559)	0.037							
		adenyl nucleotide binding (GO:0030554)	0.037							
Root Na:K	52	voltage-gated cation channel activity (GO:0022843)	0.008	LOC_Os06g14310	potassium channel AKT1, putative, expressed	AT4G32650	ATKC1,AtLKT1,KAT3,KC1	potassium channel in Arabidopsis thaliana 3	4.38	1.17E-05
		voltage-gated potassium channel activity (GO:0005249)	0.008							
		cation channel activity (GO:0005261)	0.009							
		potassium channel activity (GO:0005267)	0.009							
Root Na	42	-	-	-	-	-	-	-	-	-
Root K	40	voltage-gated cation channel activity (GO:0022843)	0.006	LOC_Os06g14310	potassium channel AKT1, putative, expressed	AT4G32650	ATKC1,AtLKT1,KAT3,KC1	potassium channel in Arabidopsis thaliana 3	-5.10	3.37E-07
		voltage-gated potassium channel activity (GO:0005249)	0.006							
		potassium channel activity (GO:0005267)	0.007							
Shoot Biomass Response	44	regulation of RNA biosynthetic process (GO:2001141)	0.025	-	-	-	-	-	-	-
		regulation of transcription, DNA-templated (GO:0006355)	0.025							
Root Biomass Response	46	purine ribonucleoside triphosphate binding (GO:0035639)	0.036	-	-	-	-	-	-	-

CHAPTER 7

CONCLUDING REMARKS

Excessive salt is a major factor that limits crop productivity. More than 800 million hectares worldwide is affected by salt, which accounts for 6% of the total land area (Munns and Tester, 2008). Rice is arguably the most important food crop worldwide, and provides a major source of calories for billions of people in tropical and subtropical developing countries. In many regions throughout Southeast Asia, rice is grown in low-lying coastal regions, which are prone to ingress by seawater. Rice is highly sensitive to salt stress, with mild salinity resulting in large reductions in yield (Zeng and Shannon, 2000; Zeng et al., 2002). Improving salt tolerance in rice is essential for ensuring food security for many individuals. However, an effective strategy must first identify traits that contribute to salt tolerance, and secondly identify loci that regulate these traits.

In this work, the overall goals were to characterize a diverse collection of cultivated rice for salt tolerance, and elucidate the genetic basis of salt tolerance-related traits. To this end, a multidisciplinary approach that combines aspects of classical plant physiology, molecular biology, computational biology, and quantitative genetics was employed. Despite the overall sensitivity to salt stress that has been previously reported by other studies, considerable variation was found within cultivated rice for many salt tolerance-related traits. The five subpopulations of cultivated rice (*indica*, *temperate japonica*, *tropical japonica*, *aromatic*, and *aus*) showed contrasting responses to moderate salt stress. In the classical phenotypic experiments described in chapter 2, the *indica* subpopulation exhibited a lower shoot Na^+ and Na^+:K^+ and higher root Na^+ and Na^+:K^+ compared to the other subpopulations. However, the image-based phenomics

approach revealed that the tropical subpopulation showed a more tolerant shoot growth response compared to the other subpopulations in rice. These results indicate that these subpopulations may harbor valuable genetic resources to improve these traits in cultivated rice.

The high density SNP information for the rice diversity panel was leveraged to characterize the genetic basis of the traits recorded from classical and image-based phenotyping. For the majority of the traits, with the exception of root Na^+ and root Na^+:K^+, many loci with very small effects were identified. These results indicate that many salt tolerance-related traits have a complex genetic architecture.

The genetic complexity of such traits presents a major challenge in further studying the molecular mechanisms and/or biological pathways that underlie these traits. However, emerging technologies for the plant science community can be used to generate a more high-resolution phenotypic data and present new opportunities to redefine traits that may be important for adaptation to saline conditions. For instance, image-based phenomics platforms can be used to nondestructively phenotype plants throughout the duration of stress, allowing time-dependent responses to be recorded. Moreover, these systems can be fitted with sensors that monitor chlorophyll health/senescence (described in chapter 5), leaf temperature, plant nutrition, and water content (Fahlgren et al., 2015). While such technologies are currently being used to provide a comprehensive phenotype, the development of new sensors will continue to improve phenotypic resolution and address more specific aspects of plant biology.

The use of intermediate phenotypes, such as metabolites, gene expression, or ions, can be used to further study the molecular mechanisms underlying complex traits. In chapter 6, transcriptional data was used to identify genes that may be involved in the regulation of salt

tolerance-related traits. Molecular phenotypes can be used to identify pathways that influence salt responses, and directly identify genes that regulate those pathways.

7.1 References

Fahlgren, N., M.A. Gehan, and I. Baxter. 2015. Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. Curr. Opin. Plant Biol. 24: 93–99.

Munns, R., and M. Tester. 2008. Mechanisms of salinity tolerance. Annu. Rev. Plant Biol. 59: 651–81.

Zeng, L., and M. Shannon. 2000. Salinity effects on seedling growth and yield components of rice. Crop Sci. 40: 996–1003Available at https://dl.sciencesocieties.org/publications/cs/abstracts/40/4/996.

Zeng, L., M.C. Shannon, and C.M. Grieve. 2002. Evaluation of salt tolerance in rice genotypes by multiple agronomic parameters. : 235–245.